1-1-2012

International Conference on MATLAB APPLICATIONS IN ENGINEERIN AND TECHNOLOGY

Prof. Srikanta Patnaik, Mentor
IRNet India, patnaik_srikanta@yahoo.co.in

Proceedings

Of

International Conference

on

MATLAB APPLICATIONS IN ENGINEERIN AND TECHNOLOGY

January 01st, 2012

Editor-in-Chief

Prof. (Dr.) Srikanta Patnaik

President, IRNet India and Chairman IIMT

Intersceince Campus, Bhubaneswar

Email: patnaik_srikanta@yahoo.co.in

Organized by:

About **ICMAET**

MATLAB is a high-level technical computing language and interactive environment for algorithm development, data visualization, data analysis, and numerical computation. In recent years, many scientist and engineers use MATLAB to solve their technical computing problems which is of course faster than with traditional programming languages, such as C, C++, and Fortran. Secondly, one can use MATLAB in a wide range of applications, including signal and image processing, communications, control design, test and measurement, financial modeling and analysis, and computational biology. This is the first international conference is being organized by Interscience Research Network (IRNet) in order to encourage our researchers who are conducting the research and bringing out the result using MATLAB.

We encourage submitting papers which are from following areas:

Application Areas of MATLAB

- Computer Science and Engineering
- Information Technology
- Electrical Engineering
- Electrical and Electronics Engineering
- Electronics and Communication Technology
- Mechanical Engineering
- Civil Engineering
- Bio-informatics and Bio Technology

The specific topics which are included but not limited to are as follows:

¤ Communications Systems ¤ Computational Biology ¤Computational Finance ¤ Control Systems ¤ Digital Signal Processing ¤ Embedded Systems ¤ FPGA Design ¤ Image and Video Processing ¤ Mechatronics ¤ Technical Computing ¤ Test and Measurement ¤ Parallel Computing ¤Math, Statistics, and Optimization ¤ Control System Design and Analysis ¤ Signal Processing and Communications ¤ Image Processing and Computer Vision ¤ Test & Measurement ¤ Computational Finance ¤ Computational Biology ¤ Application Deployment ¤ Code Generation ¤ Database Connectivity and Reporting ¤ Automotive Applications ¤ Financial Applications ¤ Aerospace Applications ¤ Portfolio Optimization ¤ Optimization Techniques ¤ Data Processing and Visualization ¤ Control Design Acceleration

EDITORIAL BOARD

Dr. B. Suryanarayana Adiga
Principal consultant
TATA Consulting Services and Visiting Professor,
IIT Powai, Mumbai

Dr. B. Anami,
Principal, K.L.E Engineering College
HUBLI, Karnataka
Email: anami_basu@hotmail.com

Prof. Ramkrushna Swain
C. V. Raman College of Engineering
At/ P.O. Bidya Nagar; P.O. Mahura; Janla
Dist: Khurda; Pin- 752054

Dr. B.V. Dhandra
Professor, Dept. of Computer Science
Gulbarga University
Gulbarga

Dr. Shivanand M. Handigund
Professor of Computer Science & Engg. and Head
M. Tech. CSE Programme & Research Centre
Bangalore Institute of Technology,
K. R. Road, V. V. Puram,
Bangalore-560 004

Dr. K.S. Sreedhara,
Professor & Chairman
Dept. of Computer Science & Engineering
University BDT College of Engineering
Davanegere University
Davengere 577004

Dr. A. Sreenivasan
Professor, PG Studies
Dayananda Sagar College of Engineering,
Kumaraswamy Layout
BANGALORE-560078

Dr. Ramaswamy
Professor, Dept. of computer Science
JAIN UNIVERSITY, Kanakapura Road
BANGALORE, Karnataka

Dr. N V Subba Reddy,
Director,
Sikkim Manipal Institute of Technology,
Majitar, Rangpo, East Sikkim

Dr. S.V. Sathyanarayana
Prof & Head
Dept. of Electronics & communication
J.N.N. College of Engineering
SHIMOGA
Karnataka

Dr. K. N. Hari Bhat,
Prof & Head,
Dept. of E & C,
Nagarjuna College of Engg., & Tech.,
Bangalore.

Dr. K.T.Veeramanju
Professor, Dept. of Electrical Engineering
J. C. Engg. College, MYSORE

First Impression : 2011

Proceedings of International Conference on **MATLAB Applications in Engineering and Technology**

DISCLAIMER

The authors are solely responsible for the contents of the papers complied in this volume. The publishers or editors do not take any responsibility for the same in any manner. Errors, if any, are purely unintentional and readers are requested to communicate such errors to the editors or publishers to avoid discrepancies in future.

ISBN: 978-93-81693-12-4

Published by :

IPM PVT. LTD., Interscience Campus
At/PO.: Kantabada, Via: Janla, Dist: Khurda, Pin- 752054
Publisher's Website : www.interscience.in
E-mail: ipm.bbsr@gmail.com

Typeset & Printed by :
IPM PVT. LTD.

TABLE OF CONTENTS

Sl. No.	Topic	Page No.
	Editorial - *Prof. (Dr.) Srikanta Patnaik*	
1	**Study of Embedded Multiprocessor Architecture Using FPGA** - *Sumedh. S. Jadhav & C. N. Bhoyar*	01-06
2	**Iris Movement Controlled Wheel Chair** - *Lekha Das & Uday Chhatre*	07-09
3	**Optimal Power Flow Analysis based Available Transfer Capability Calculations in Deregulated Power System** - *Sunil kumar Chava, J.Amarnath & P.S Subramanyam*	10-18
4	**Differential Diagnosis of Dementia Using Slantlet Transform** - *M.M.Patil & A.R.Yardi*	19-23
5	**Modeling of Novel Control Strategy used in Active Harmonic Filter for harmonic mitigation and reactive power compensation** - *Vivek Vora & Ramanujam Cadatur*	24-28
6	**AN APPLICATION OF MATLAB/SIMULINK FOR SPEED CONTROL OF DC SERIES MOTOR USING BUCK CHOPPER** - *Joy Bandhyapadhyay, Arghya Mukherjee & Suman Ghosh*	29-33
7	**CLUSTER : A Matlab GUI Package for Data Clustering** **Sudip Poddar, Anirban Mukhopadhayay**	34-40
8	**A Novel Multi-Planar Band Structure Algorithm : Planar Region Matrix Approach** - *Deepak Bharti & Abhijit R. Asati*	41-48
9	**Clutter Reduction in Ground Penetrating Radar B-Scan Imagery by Gaussian filtering for Improved Landmine Recognition** - *P. Iswerya*	49-54
10	**Error Performance of Turbo Codes in AWGN Channels** - *P. Surendra Kumar, B.Chandra Mohan & T.Krishna Chaitanya*	55-60
11	**Signal Processing and Fault Detection in Induction Motors** - *Shrawani Badoniya, Sulochana Wadhwani & A.K.Wadhwani*	61-66

12 **Implimentation of A Parallel IEEE P754 Decimal Floating-Point Multiplier Using Verilog** 67-72

- *Vrushali Ailawar*

13 **All Digital Phase Locked Loop (ADPLL) For High Performance Microprocessors** 73-77

- *P. Bosebabu & K. Rakesh*

14 **Comparative Analysis of Image Compression Using Image Interpolation And RICE Codes With Quadtree Encoding** 78-84

- *Vaibhav Jain, Nitin Jain, Ashish Parikh*

15 **Face Verification Across Age Progression Using Discriminative Methods** 85-92

- *Piyush D. Hegu, Ashwin W. Motikar & Pradhnya J. Suryawanshi*

16 **Dynamic Stability Improvement With Combined Fast Acting Automatic Voltage Regulator (AVR) And Automatic Load Frequency Control (ALFC) Loops** 93-96

- *Venu Yarlagadda , K. R. M. Rao, P. Sai Saroja, P. Avinash & P. Nikhilesh*

17 **Design And Implementation For Identification Of Moisture Content In Cotton Bale By Microwave Imaging** 97-100

- *Ashish V. Saywan*

18 **Application of Virtual Routers in Wireless Communication** 101-104

- *Ashwin W. Motikar & Narendra Narole*

19 **Speaker Recognition Using K-Means Algorithm** 105-108

- *Kavita Bhangale & Shoba Krishnan*

20 **Implementation of Modified Architecture for Adaptive Viterbi Decoder** 109-114

- *Pallavi R. More & Divya M. Meshram*

21 **Non Linear Controller for DC-DC Buck Converter** 115-119

- *T. Chinthu & Poorani Shivkumar*

22 **Speech Recognition using SOM and Actuation via Network in Matlab** 120-124

- *Subash Chandar. A, Suriyanarayanan. S & Manikandan. M*

23 **Information Hiding In Color Images** 125-129

- *R. S. Mahamune & P. J. Suryawanshi*

24 **Nonlinear System Identification Using A Novel Immune Artificial Fish Swarm Algorithm** 130-134

- *Ritesh Kumar & Nishant Sahay*

25 **Design of Fault-Tolerant Memory with Fault Secure Encoder And Decoder** 135-141

- *Sunitha R, Supritha G. S. & Ashwini*

26 **Gnome Sorted Hessian Percolation Model For Automated Visual Quality Inspection of Manufactured Metallic Products** 142-148

- *Sarin CR ME*

27 **The Challenges of Effectively Anonymizing Network Data** 149-156

- *Ch. Himabindu, Y. Chitti Babu & P. Harini*

INTERNATIONAL CONFERENCE ON MATLAB APPLICATIONS IN ENGINEERING AND TECHNOLOGY

Almost every researcher will not deny the fact that Computational Complexity is increasing with the advanced trend multidisciplinary approach to the field of Computation. In many instances the intricacies are not being solved through programming languages. MATLAB is an interactive environment designed to perform scientific and engineering calculations and to create computer simulations. MATLAB is a customized computing solution to the engineering researchers. MATLAB is a high-level language and interactive environment that enables you to perform computationally intensive tasks faster than with traditional programming languages such as C, C++, and FORTRAN. MATLAB is attributed to the pioneering work of Cleve Moler, head of Computer Science Department, University of New Mexico during 1970s. The development of the field is credited to the works of MathWorks during 1984.

MATLAB was first adopted by researchers and practitioners in control engineering, Little's specialty, but quickly spread to many other domains. It is now also used in education, in particular the teaching of linear algebra and numerical analysis, and is popular amongst scientists involved in image processing. MATLAB can interface with other languages also. MATLAB can call functions and subroutines written in the C programming language or FORTRAN. A wrapper function is created allowing MATLAB data types to be passed and returned. The dynamically loadable object files created by compiling such functions are termed "MEX-files" (for MATLAB **ex**ecutable). Since last September 2011 MATLAB advance version 7.13 is available for users. Simulink is integrated with MATLAB as MATLAB/Simulink, i.e., data can be easily transferred between the programs. MATLAB is supported in UNIX, Macintosh, and Windows environments. This way, Simulink is an interactive environment for modeling, analyzing, and simulating a wide variety of dynamic systems.

The use of MATLAB is actually increasing in a large number of fields, by combining with other toolboxes, e.g., optimization toolbox, identification toolbox, and others. The MathWorks Inc. periodically updates MATLAB and Simulink, providing more and more advanced software. MATLAB handles numerical calculations and high-quality graphics, provides a convenient interface to built-in state-of-the-art subroutine libraries, and incorporates a high-level programming language. Nowadays, the MATLAB/Simulink package is the world's leading mathematical computing software for engineers and scientists in industry and education.

Due to the large number of models and/or toolboxes, there is still some work or coordination to be done to ensure compatibility between the available tools. Inputs and outputs of different models are to-date defined by each modeler, a connection between models from two different toolboxes can thus take some time. This should be normalized in the future in order to allow a fast integration of new models from other toolboxes. The widespread use of these tools is reflected by ever-increasing number of books based on the MathWorks Inc. products, with theory, real-world examples, and exercises.

The conference is designed to stimulate the young minds including Research Scholars, Academicians, and Practitioners to contribute their ideas, thoughts and nobility in these two integrated disciplines. Even a fraction of active participation deeply influences the magnanimity of this international event. I must acknowledge your response to this conference. I ought to convey that this conference is only a little step towards knowledge, network and relationship

I express best wishes to all the paper presenters. I extend my heart full thanks to the reviewers, editorial board members, programme committee members of the conference. If situations prevail in favor we will take the glory of organizing the second conference of this kind during this period next year.

Editor-in-Chief

Prof. (Dr.) Srikanta Patnaik
President IRNet and Chairman, I.I.M.T., Bhubaneswar
Intersceince Campus,
At/Po.: Kantabada, Via-Janla, Dist-Khurda
Bhubaneswar, Pin:752024. Orissa, INDIA

Study of Embedded Multiprocessor Architecture Using FPGA

Sumedh. S. Jadhav & C. N. Bhoyar

Dept of Electronics Engg, Nagpur,India
Dept of Electronics Engg., Priyadarshini College of Engineering, Nagpur,India
E-mail : sumedh_jadhav@rediffmail.com, cnbhoyar @yahoo.com

***Abstract* -** Field programmable gate arrays (FPGAs) have the potential to accelerate scientific computing applications due to their highly parallel architecture However, for programming these architectures efficiently, hardware description languages (HDL), such as Verilog or VHDL, have to be used. Embedded multiprocessor design presents challenges and opportunities that stem from task coarse granularity and the large number of inputs and outputs for each task. We have therefore designed a new architecture called embedded concurrent computing (ECC), which is implementing on FPGA chip using VHDL. The design methodology is expected to allow scalable embedded multiprocessors for system expansion. In recent decades, two forces have driven the increase of the processor performance: Advances in very large-scale integration (VLSI) technology and Micro architectural enhancements. Therefore, we aim to design the full architecture of an embedded processor for realistic to perform arithmetic, logical, shifting and branching operations. We will be synthesize and evaluated the embedded system based on Xilinx environment. Processor performance is going to be improving through clock speed increases and the clock speed increases and the exploitation of instruction- level parallelism. We will be designing embedded multiprocessor based on Xilinx environment.

Keywords*—FPGA based embedded system design; real time processor; computer architecture.*

I. INTRODUCTION

In recent decades, two forces have driven the increase of the processor performance: Firstly, advances in very large-scale integration (VLSI) technology and secondly micro architectural enhancements [1].

Processor Performance has been improve through clock speed Increases and the exploitation of instruction-level Parallelism. While transistor counts continue to increase, recent attempts to achieve even more significant increase in single-core performance have brought diminishing returns [2, 3]. In response, architects are building chips With multiple energy-efficient processing cores instead of investing the whole transistor count into a single, complex, and power-inefficient core [3, 4]. Modern embedded systems are design as systems-on a-chip (SoC) that incorporate single chip multiple Programmable cores ranging from single chip multiple programmable cores ranging from processors to custom designed accelerators. This paradigm allows the reuse of pre-designed cores, simplifying the design of billion transistor chips, and amortizing costs. In the past few years, parallel-programmable SoC (PPSoC)have Successful PPSoC are high-performance embedded multiprocessors such as the STI Cell [3] .They are dubbed single-chip heterogeneous multiprocessors (SCHMs) because they have a dedicated processor that coordinates the rest of the processing units. A multiprocessor design with SoC like integration of less-efficient, general-purpose processor cores with more efficient special-purpose helper engines is project to be the next step in computer evolution [5].

First, we aim to design the full architecture of an embedded processor for realistic throughput. We used FPGA technology not only for architectural exploration but also as our target deployment platform because we believe that this approach is best for validating the feasibility of an efficient hardware implementation. Nevertheless, when implemented, we attempted to ensure that some parts of the architecture and the corresponding Register Transfer Level (RTL) Implimentation are platformagnostic.

This architecture of the embedded processor resembles a superscalar pipeline, including the fetch, decode, rename, and dispatch units as parts of the in-order front-end. The out of-order execution core contains the task queue, dynamic scheduler; execute unit, and physical register file. The in order back-end is comprised of only the retire unit. The embedded architecture will be implementing using the help of RTL descriptions in System VHDL.

We will integrate the embedded processor with a shared memory system, synthesized this system on an FPGA environment, and performed several experiments using realistic benchmarks. The methodology to design

and implement a microprocessor or multiprocessors is presented. To illustrate it with high detail and in a useful way, how to design the most complex practical session is shown. In most cases, computer architecture has been taught with software simulators [1], [2]. These simulators are useful to show: internal values in registers, memory accesses, cache fails, etc. However, the structure of the microprocessor is not visible.

In this work, a methodology for easy design and real Implementation of microprocessors is proposed, in order to provide students with a user-friendly tool. Simple designs of microprocessors are exposed to the students at the beginning, rising the complexity gradually toward a final design with two processors integrated in an FPGA; each of which has an independent memory system, and are intercommunicated with a unidirectional serial channel.

II. MULTIPROCESSOR

Multiprocessor system consists of two or more Connect processors that are capable of communicating. This can be done on a single chip where the processors are connected typically by either a bus. Alternatively, the multiprocessor system can be in more than one chip, typically connected by some type of bus, and each chip can then be a multiprocessor system. A third option is a multiprocessor system working with more than one computer connected by a network, in which each Computer can contain more than one chip, and each chip can contain more than one processor.

A parallel system is presented with more than one task, known as threads. It is important to spread the workload over the entire processor, keeping the difference in idle time as low as possible. To do this, it is important to coordinate the work and workload between the processors. Here, it is especially crucial to consider whether or not some processors are special-purpose IP cores. To keep a system with N processors effective, it has to work with N or more threads so that each processor constantly has something to do. Furthermore, it is necessary for the processors to be able to communicate with each other, usually via a shared memory, where values that other processors can use are stored. This introduces the new problem of thread safety. When thread safety is violated, two processors (working threads) access the same value at the same time. Some methods for restricting access to shared resources are necessary. These methods are known as thread safety or synchronization. Moreover, it is necessary for each processor to have some private memory, where the processor does not have to think about thread safety to speed up the processor. As an example, each processor needs to have a private stack. The benefits of having a multiprocessor are as follows:

1. Faster calculations are made possible.
2. A more responsive system is created.
3. Different processors can be utilized for different

Tasks. In the future, we expect thread and process parallelism to become widespread for two reasons: the nature of the Applications and the nature of the operating system. Researchers have therefore proposed two alternatives Micro architectures that exploit multiple threads of Control: simultaneous multithreading (SMT) and chip multiprocessors (CMP). Chip multiprocessors (CMPs) use relatively simple.

Single-thread processor cores that exploit only moderate amounts of parallelism within any one thread, while executing multiple threads in parallel across multiple processor cores. Wide-issue superscalar processors exploit instruction level parallelism (ILP) by executing multiple instructions from a single program in a single cycle. Multiprocessors (MP) exploit thread-level parallelism (TLP) by executing different threads in parallel on Different processors.

III. SOFTWARE TOOL

The Xilinx Platform Studio (XPS) is used to design Micro Blaze processors. XPS is a graphical IDE for developing and debugging hardware and software. XPS simplifies the procedure to the users, allowing them to select, interconnect, and configure components of the final system. Dealing with this activity, the student learns to add processors and peripherals, to connect them through buses, to determine the processor memory extension and allocation, to define and connect internal and external ports, and to customize the configuration parameters of the components. Once the hardware platform is built, the students learn many concepts about the software layer, such as: assigning drivers to Peripherals, including libraries, selecting the operative system (OS), defining processor and drivers parameters, assigning interruption drivers, establishing OS and libraries parameters.

An embedded system performed with XPS can be Summarized as a conjunction of a Hardware Platform (HWP) and a Software Platform (SWP), each defined separately.

A. HARDWARE PLATFORM

The HWP is described in the Microprocessor Hardware Specification (MHS) file; it contains the description of the system architecture, the memory map and the configuration parameters. HWP can be defined as one or more processors connected to one or more peripherals through one or more buses. The definition of the activity follows this sequence:

- To add processors and peripherals.
- To connect them through buses.
- To determine the processor memory allocation.
- To define and connect internal and external ports.
- To customize the configuration parameters of the Components.

B. THE SOFTWARE PLATFORM

The SWP is described in the Microprocessor Software Specification (MSS) file; it contains the description of drivers, component libraries, configuration parameters, standard input/output devices, interruption routines and other software features. The sequence of activities needed to define the SWP is the following:

- To assign drivers to peripherals.
- To assign interruption drivers.
- To establish OS and libraries' parameters.
- To assign Input/output port.
- To assign timers.
- To establish components parameters

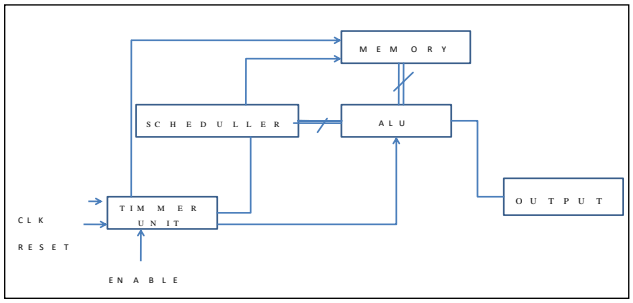


Figure 1. Block diagram of Multiprocessor

C. SYSTEM MEMORY

A system that complies with the MP specification uses the standard AT memory architecture. All memory is allocated for system memory with the exception of addresses 0A_0000h through 0F_FFFFh and 0FFFE_0000h through 0FFFF_FFFFh, which are reserved for I/O devices and the BIOS.

Compared to a uniprocessor system, a symmetric multiprocessor system imposes a high demand for memory bus bandwidth. The demand is proportional to the number of processors on the memory bus. To reduce memory bus bandwidth limitations, an implementation of this specification should use a secondary cache that has high-performance features, such as a write-back update policy and a snooping cache-consistency protocol. A secondary cache can push the scalability limit upward by reducing bus traffic and increasing bus bandwidth.

In this form of multiprocessing, each background process renders its own frame and runs on a separate processor core (CPU). The number of processes used to render multiple frames simultaneously is never more than the number of processors. The number of background processes that can run on your computer also depends on the total amount of installed system RAM and the amount of RAM that is assigned to the After Effects application.

D. SCHEDULLER

- To assign drivers to peripherals.
- To assign interruption drivers.
- To establish OS and libraries' parameters.
- To assign drivers to performs specific tasks.
- To support multiprogramming
- Large numbers of independent processes
- Simplified administration

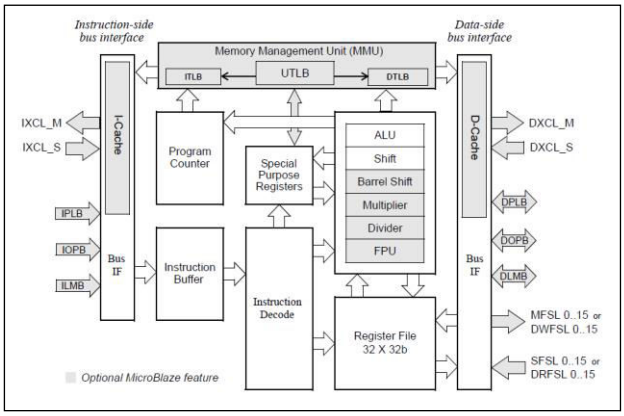


Figure 2. Proposed architecture of multiprocessor architecture.

IV. THE MICROBLAZE MULTIPROCESSOR PROCESSOR

Micro Blaze Multiprocessor is a 32-bit specific purpose processor Developed by Xilinx in VHDL. It can be parameterized using XPS to obtain an *à-la-carte* processor. It is a RISC processor, structured as Harvard architecture with separated data and instruction interfaces. Micro Blaze components are divided into two main groups depending on their configurability as shown in Fig.1. Some fixed feature components are:

- 32 general purpose registers sized 32-bit each.

- Instructions with 32 bits word-sized, with 3 operands and 2 addressing modes.
- 32 bits address bus.
- 3-stage Pipeline.

Some of the most important configurable options are:

- An interface with OPB (On-chip Peripheral Bus) data bus.
- An interface with OPB instruction bus.
- An interface with LMB (Local Memory Bus) data bus.
- An interface with LMB instruction bus.
- Instruction cache.
- To include EDK libraries.
- To select the operative system (OS).
- To define processor and drivers' parameters.
- Data cache.
- 8 Fast Simplex Link (FSL bus) Interfaces.
- Cache Link bus support.
- Hardware exception support.
- Floating Point Unit (FPU).

The suggested core embedded processor contains a dual-issue, superscalar, pipelined processing unit, along with the other functional elements required to Implement embedded SoC solutions. This other Functions include memory management and timers.

V. PRACTICAL DESIGNS

Practical sessions introduce gradual learning, allowing the fast design based on previous sessions. Essential problems in hardware programming will be raised:

- HyperTerminal serial communication.
- Using IO ports.
- Memory controller.
- Interruption routines and priority.

A. *PIPELINE ARCHITECTURE*

Micro Blaze Multiprocessor instruction execution is pipelined. For most instructions, each stage takes one clock cycle to complete. Consequently, the number of clock cycles necessary for a specific instruction to complete is equal to the number of pipeline stages, and one instruction is completed on every cycle. A few instructions require multiple clock cycles in the execute stage to complete. This is achieved by stalling the pipeline.

When executing from slower memory, instruction fetches may take multiple cycles. This additional latency directly affects the efficiency of the pipeline. MicroBlaze Multiprocessor implements an instruction prefetch buffer that reduces the impact of such multi-cycle instruction memory latency. While the pipeline is stalled by a multi-cycle instruction in the execution stage, the prefetch buffer continues to load sequential instructions. When the pipeline resumes execution, the fetch stage can load new instructions directly from the prefetch buffer instead of waiting for the instruction memory access to complete.

3-STAGE PIPELINING

When area optimization is enabled, the pipeline is divided into three stages to minimize hardware cost: Fetch, Decode, and Execute.

	cycle 1	cycle 2	cycle 3	cycle4	cycle5	cycle6	cycle7
instruction 1	Fetch	Decode	Execute				
instruction 2		Fetch	Decode	Execute	Execute	Execute	
instruction 3			Fetch	Decode	Stall	Stall	Execute

B. *BRANCHES*

Normally the instructions in the fetch and decode stages (as well as prefetch buffer) are flushed when executing a taken branch. The fetch pipeline stage is then reloaded with a new instruction from the calculated branch address. A taken branch in MicroBlaze takes three clock cycles to execute, two of which are required for refilling the pipeline. To reduce this latency overhead, MicroBlaze Multiprocessor supports branches with delay slots. When executing a taken branch with delay slot, only the fetch pipeline stage in MicroBlaze is flushed. The instruction in the decode stage (branch delay slot) is allowed to complete. This technique effectively reduces the branch penalty from two clock cycles to one.

C. *MEMORY ARCHITECTURE*

MicroBlaze is implemented with Harvard memory architecture; instruction and data accesses are done in separate address spaces. Each address space has a 32-bit range (that is, handles up to 4-GB of instructions and data memory respectively). The instruction and data memory ranges can be made to overlap by mapping them both to the same physical memory. The latter is useful for software debugging.

MicroBlaze does not separate data accesses to I/O and memory (it uses memory mapped I/O). The

processor has up to three interfaces for memory accesses:

- Local Memory Bus (LMB)
- Processor Local Bus (PLB) or On-Chip Peripheral Bus (OPB)
- Xilinx Cache Link (XCL) The LMB memory address range must not overlap with PLB, OPB or XCL ranges.

VI. BIPROCESSOR SYSTEM DESIGN

The last and most complex practical session is the design and implementation of a biprocessor. A computational system composed of two Micro Blazes will be designed. Both Micro Blazes will be interconnected using message-passing protocol. Each Micro Blaze has its own non-shared memory for Instructions and data. In the Fig. 3 a diagram with the structure of the design is shown. In it, the buses and components used have been detailed. It also includes how they are interconnected At first, following the logical sequence exposed previously, a HWP will be created. This HWP will include the configuration of the components and buses, their interconnection, the memory map, ports and other parameters. In the following subsection, the steps needed to configure the system will be described. The parameters shown in this section depends on the FPGA chip, in this case the Spartan 3 board [11].

A. HARDWARE PLATFORM SPECIFICATIONS

This stage is described in the MHS file. Following, the Components specified in the structure of the system are Enumerated:

- Two Micro Blaze processors.
- Two on-chip RAM memory blocks (BRAM), one for Each processor.
- One UART.
- One OPB bus, to connect the UART with the slave Processor.
- Two LMB buses to communicate each processor with their respective data memory controller; and another two LMB buses to interconnect the processors with their instruction memory controller.
- One FSL channel to intercommunicate each processor with the other.

After that, the interconnection of buses and components is defined. The connection of the memory ports are also set at this point. The student has to specify in the connection matrix which components are linked to which buses and with which kind of connection.

In the exposed case, four LMB buses are needed to access local memory, two for each Micro Blaze, because each processor has its own memory subsystem.

Also, one FSL channel which connects both processors. Each BRAM has been designed with 4 different ports. Each Micro Blaze reaches its memory block through two different interfaces (instructions and data). After that, it is necessary to map the components inside the Configuration memory of the processors. XPS provides a functionality which is able to compute automatically a valid configuration memory map for a monoprocessor system structure. However, as the system proposed is a biprocessor one, this functionality cannot Each Micro Blaze looks for the first instruction in its Program at memory address 0x0. be used. The next step is to define the internal and external ports.

Most of the internal ones are configured by XPS with default settings. It is also necessary to define and to connect some of the internal ports to make the system works: those ports related to the reset and clock signals must be forwarded to all of the subsystems and components. Four external ports are mandatory: clock, reset, UART in and UART out. With these ports, the student sends commands and synchronization Information to the system. Finally, the components are configured. The parameters for each component and their meaning are described thoroughly in the documentation included in the XPS platform.

Particularly, Micro Blaze includes a parameter which selects the amount of FSL interfaces used. Thus, both processors have to set this configuration value to one to allow the communication between them. The configuration of this parameter is done by changing C_FSL_LINKS. This parameter has to be set to a numerical value, representing the amount of FSL interfaces to be included in the core.

Another interesting configuration to be mentioned is the UART operational configuration. The student has to determine the operational frequency, the application of the parity bit checking, working bauds, etc. A valid set of parameters for the UART and Micro Blaze are the following:

1) UART parameters.

a) C_CLK_FREQ = 50_000_000. Set the frequency of the OPB bus, connected to the UART. It has to coincide with the operational system speed.

b) C_BAUDRATE = 19200. Set the bauds for the UART. The terminal used to receive characters has to be configured at the same baud rate.

c) C_USE_PARITY = 0. Set whether the UART should work with parity bit or not.

VII. SOFTWARE AND HARDWARE REQUIREMENT

For Software simulation I will prefer MODELSIM and for synthesis I will be prefer XILINX. Hardware requirement is SPARTAN-3.

IX. RESULT VERIFICATION AND ANALYSIS

Observe the required result like arithmetic, logical, branching and shifting.

ACKNOWLEDGEMENTS

Authors wish to remark the great task carried out by the Xilinx and Modelsim user guide; and the authors wish to thank C.N.Bhoyar for his contribution in the design process.

REFERENCES

[1] John L. Hennessy and David A. Patterson. Computer architecture: a quantitative approach. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, fourth edition 2007.

[2] Doug Burger and James R. Goodman. Billion-Transistor Architectures: There and Back Again. IEEE Computer, 37(3):22– 28, 2004.

[3] David Geer. Industry Trends: Chip Makers Turn to Multicore Processors. Computer, 38(5):11–13, May,2005.

[4] AMD Corporation. Multi-core processors: the next revolution in computing White paper, 2005.

[5] B. Ackland, A. Anesko, D. Brinthaupt, S.J. Daubert, A. Kalavade, J. Knobloch, E. Micca, M. Moturi, C.J. Nicol, J.H. O'Neill, J. Othmer, E. Sackinger, K.J. Singh, J. Sweet, C.J. Terman, and J. Williams. A Single-chip, 1.6-billion, 16-b MAC/s Multiprocessor IEEE Journal of, 35(3):412–424, Mar 2000.

[6] Asawaree Kalavade, Joe Othmer, Bryan Ackland, and K. J. Singh. Software environment for a multiprocessor DSP. In DAC '99: Proceedings of the 36th ACM/IEEE conference on Design automation, pages 827–830, New York, NY, USA, 1999. ACM..

[7] V. Sklyarov, and I. Skliarova. "Teaching Reconfigurable Systems:Methods, Tools, Tutorials, and Projects," *IEEE Trans. on Education,vol. 48, no. 2, 290–300,* 2005.

[8] OpenSPARC http:// www.opensparc.net/ edu/university-program.html. Last accessed on 8th November 2009.

[9] "Platform Studio User Guide," Application notes, Xilinx, 2005.

[10] "Microblaze Processor Reference Guide," Application notes, Xilinx, 2005.

[11] "Embedded System Tools Reference Manual," Application notes, Xilinx, 2008.

[12] "OS and Libraries Document Collection," Xilinx, Application notes, September 2007.

[13] Spartan-3 Board. http://www.digilentinc.com/ . Last accessed on 30th October 2009.

◈◈◈

Iris Movement Controlled Wheel Chair

Lekha Das & Uday Chhatre
K J Somaiya Collegeof Engineering, Mumbai-77
E-mail : lekhadas@engg.somaiya.edu,udaychhatre@engg.somaiya.edu

Abstract **-** Physically multiple disabled needs assistance. If the person is paralytic below the neck, there should be some aid for him/her without a permanent assistance. This can be achieved by using a wheelchair controlled by the measurement of eyeball movement. The design strategy is making a head-gear for the control of movement of wheelchair. Research is ongoing using Electrooculographic (EOG) Electrodes mounted on the face of the patient or web cam mounted on the head for the measurement of eyeball movements. The proposed system is implemented using Infra Red (IR) dual sensors to measure movements of iris to control the movement of wheelchair. The project is aimed for spastic children having cerebral palsy with no upper dexterity and partially lower.

Keywords- *differential mode motor, EOG, Infra Red sensor, saccadic movements.*

I. INTRODUCTION

One of the most developing researches in engineering that utilizes the extensive research in medicine is biomedical engineering. Communication between humans seem usually much simple than the one involves humans and machines. This difficulty increases when a person is disabled. Especially this kind of people has more to gain by assisting a machine in everyday life. Vision is one of our most valued senses and during the course of each day our eyes are constantly moving. There are three antagonistic muscle pairs, which relax and contract in order to induce eye movement attached to the globe of the eye. [1] These pairs of muscles are responsible for horizontal, vertical and torsional (clockwise and counter clockwise) movement. The human input is the electronic signals produced by moving eyes. There are different ways to measure the signals produced by eyeball movements. [2]

The main aim of this work is to provide a better independent life to those with multiple disabilities. Eyeball movement is used to control the movement of wheelchair the wheelchair equipped with EOG electrodes also measure the eyeball movements. The most important disadvantages related to the fact that the corneo-retinal potential is not fixed but has been found to vary, and to be affected by light, fatigue, and other qualities. Consequently, there is a need for frequent calibration. Additional difficulties arise owing to muscle artifacts and the basic nonlinearity of the method. Another method uses web cam which is bulky and costly.

This system will also use eye blinking for start and stop of the movement of the wheelchair. It will sense any obstacle in the path and alarm the user to avoid collision. The eye movement signals are transferred to wheelchair using electrical signals which are used to drive the motors attached to the wheels. Differential mode motors are used.

II. METHODS FOR EYE TRACKING

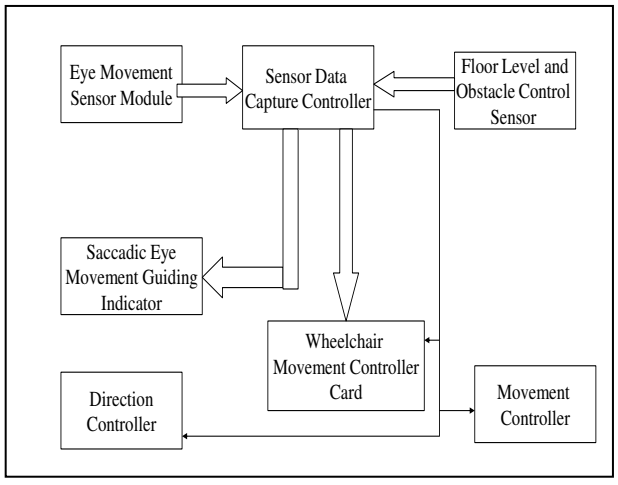
The previous researches show the use of EOG or a web cam based system for the measurement of eyeball movement and using it for the movement of wheelchair of a disabled person. 1 .Electrooculography (EOG) is a technique for measuring the resting potential of the retina. EOG is the study and interpretation of Electro-Encephalograms made by moving the eyes a constant distance between two fixed points and generates the potential .The potential can be considered as a steady electrical dipole. The resulting signal is called the electrooculogram.

Figure 1.Electrode Placement in EOG

The measurement of horizontal eye movements are done by the placement of a pair of electrodes at the outside of the left and right eye (outer canthi) as shown in figure 1. With the eye at rest the electrodes are effectively at the same potential and no voltage is recorded. The rotation of the eye to the right results in a difference of potential, with the electrode in the direction of movement (i.e., the right canthus) becoming positive relative to the second electrode. The limitations of EOG include CRP(Corneo-Retinal Potential) is not fixed, but vary due to light, fatigue etc., difficulties arising owing to muscle artifacts, need for calibration and recalibration. Moreover a skilled assistant is required for the placement of electrodes and assistance.

2. The second method used for eye tracking is using a web-cam. The eye motion tracking hardware includes a USB web camera which is mounted on a cap worn by the user. This camera is adjusted so that it lies in front of one of the eye of user. The camera has inbuilt light source, so that it can capture bright images if darkness appears under the cap. The drivers of the camera are installed in a PC to which the camera is plugged in. The software module for image processing works on three different modules: video capturing, frame extraction and pixel color detection. The limitation of this method is its bulkiness and cost.

This research uses sensors fitted on a goggle worn by the user to detect the eyeball movement to control the wheel chair.



The wheel chair control consists of two modules, the sensor module and wheel chair control module. The sensor module has LED based direction annunciation guiding mechanism (figure 3), Eye movement detection using IR (Infra Red) sensor pairs (figure 4), saccadic movement stabilization using IR sensor pairs (figure 5), eye blink detection using IR sensor pairs module (figure 6).

The collapsible direction guiding mechanism is equipped with three different colored LEDs located at 8-10 inches away from the head gear assisting the patient to fix the location for eye movements. This improves smooth turning movement mechanism. At start the Light Emitting Diodes (LEDs) (both green and red) are blinking. When iris is fixed on right side green LED on right is ON and other two are OFF. When fixed on left side green LED on left is ON and other two are OFF.

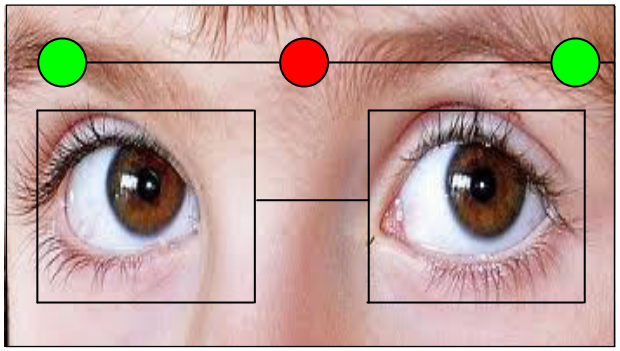


Figure 3. Direction Guiding Mechanism

The eye movement detection sensors find the location of iris in the horizontal plane of the sensor array. Thus it controls the direction of wheel chair movement

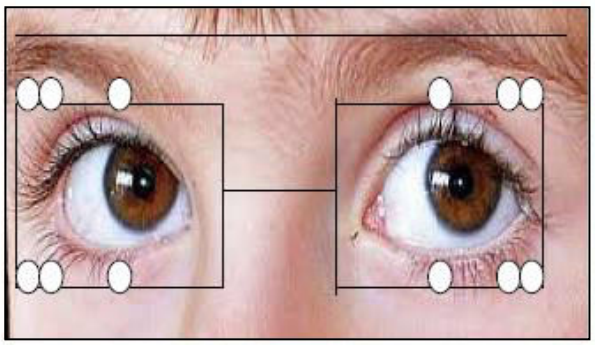


Figure 4. Direction Control

Saccades are abrupt movements of the eye that change the point of fixation (looking from one point to another point). Saccades can be voluntary or involuntary. An example of voluntary movement would be when an individual chooses to look at one side of a room and then the other. An example of involuntary movement would be the unconscious motions of the eye that occur during sleep (known as rapid eye movement or REM). The saccadic movement [5] has to be stabilized. This is achieved by using IR detectors mounted on the goggle as shown in Figure 5. Saccadic movements of eyes are considered as a bouncing. De-

bouncing is attained by the corner (left and right) IR sensors. If the values are "1-1" saccadic movements are fixed (stable eye position).

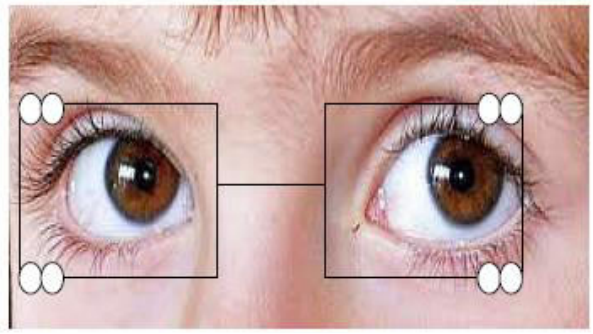


Figure 5. Saccadic Movement Fixation

The blinking of the eyes will start or stop the wheelchair. Eye blinking of 1se is used for starting and stopping of the system. Too short (natural blink) or too long blinks (drowsiness) are considered as invalid blinks.

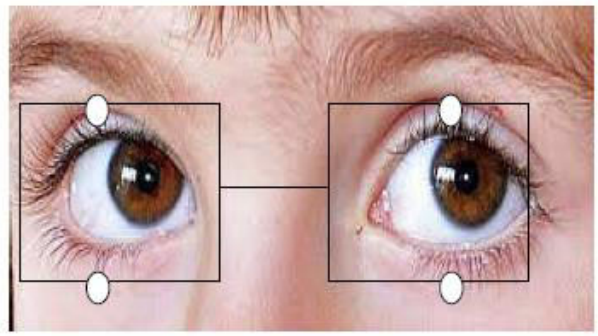


Figure 6. Blink Detection

Two level sensors LS1 and LS2 to detect the floor leveling as shown in the table 1. As soon as any of these obstacles occurs the wheelchair comes to an emergency halt. An alarm is also provided to indicate the emergency halt, the alarm can be sound or a vibration mode if the subject is hearing impaired.

LS2	LS1	Level Status of Floor
1	1	**Leveled Floor**
1	0	**Up(Staircase)**
0	1	**Down(Slope)**
0	**0**	**Don't Care**

Table 1. Level Sensor Truth Table.

Sensor module keeps on checking left/right movement of the eyes after the saccadic movement fixation. This will keep on sending incremental pulse to the wheelchair control module. With respect to that the wheelchair will take ten degree incremental turn [4] for each incremental pulse (received from the sensor module). The anti turn movement will start when the subject looks straight and keeps looking straight. The direction of the wheelchair as per the sensor data is driven by differential Pulse Width Modulator (PWM) [3] drive for left and right motors of wheel chair.

The Reflective sensor OPB 706B is used to detect the location of colored iris

IV CONCLUSIONS

The proposed system is easy to operate by the user without any skilled assistance. Compared to the previous researches of eye movement detection, this method is more user friendly. With trained subject, it can have maximum utility.

REFERENCES

[1] R. Barea, L. Boquete, M. Mazo, and E. López. "Guidance of wheelchair using electrooculography". Proceedings of the 3rd IMACS International Multi-conference on Circuits, Systems, Communications and Computers (CSCC'99). Greece July 1999

[2] M. S. Reddy, A.Sammaiah, B. Narsimha and K. S. Rao International Conference on Multimedia and Signal Processing,2011.

[3] P. Britto,Indumathi.J,Sudesh Sivarasu, and Lazar Mathew, "Automation of Wheel chair Using Ultrasonic and Body Kinematics", NCCI 2010 – National Conference on Computational Instrumentation CSIO Chandigarh, INDIA, 19-20 March 2010 .

[4] M.A.Fkirin, S.Badawy, A.S.El-Sherbeny Driving a DC Motor by Numerically Manipulated Eye Signal Captured by EOG.The Online Journal on Electronics and Electrical Engineering (OJEEE) Vol. (2) – No. (1)

[5] Available on line : www.medicin.mcgill.ca/ phisio/vlab/other_exps/EOG/ scccades_n.htm

◈◈◈

Optimal Power Flow Analysis based Available Transfer Capability Calculations in Deregulated Power System

Sunil kumar Chava[1], J.Amarnath[2] & P.S Subramanyam[3]
[1]Dept. of EEE, CVRCE, Hyderabad, [2]Dept. of EEE, JNTUH, Hyderabad, [3]Dept. of EEE, VBIT, Hyderabad
E-mail : ursunil@rediffmail.com, amarnathjinka@yahoo.com, subramanyamps@gmail.com

Abstract - In this paper, important factors that may affect generating companies' profit margins through wholesale electricity trading are discussed. These factors include generators' efficiencies and capabilities, types of generators owned, fuel costs, transmission losses and settling price variation. It demonstrates how proper analysis of these factors using the solutions of Optimal Power Flow (OPF), can allow companies to maximize overall revenue. And through this OPF analysis, companies will be able to determine, for example, which generators are most economical to run, best locations for generators to be situated at, and also the scheduling of generators as demand changes throughout the day. It illustrates how solutions of OPF can be used to maximize companies' revenue under different scenarios. And is also extended to computation of Available Transfer Capability (ATC) is very important to the transmission system security and market forecasting. From these results it is observed that how crucial it is for companies to plan their daily operations and is certainly useful in an online environment of deregulated power system. In this paper above tasks are demonstrated on IEEE 26-bus system and IEEE 118-bus system and results have been presented and analyzed.

Keywords- OPF, ATC, Electricity Market, Spot Price..

I. INTRODUCTION

In the past, the electricity industry was government-controlled and also monopolistic. However over the past decade, the industry in many countries including Australia had undergone significant changes and was restructuring for a free market, also known as deregulation. This led to a competitive market whereby customers are able to choose their electricity supply from a number of generating companies and retailers. In this deregulated market, it is essential for generating companies to plan their operations efficiently, so as to minimize operating costs while maximizing their profit margins.

There are many factors involved in the successful operation of a power system. The system is expected to have power instantaneously and continuously available to meet power demands. It is also expected that the voltage supplied will be maintained at or near the nominal rated value. Not only must the demands be met at all times, the public and employees should not be placed in hazard by operations of the system. At the same time proper operating procedures must be observed to avoid damage to equipment or other facilities of the system. All of these operating requirements must be achieved simultaneously [1].

Other than those mentioned above, one of the most important factors is the operating cost. Generation and distribution of power must be accomplished at minimum cost but with maximum efficiency. This involves the real and reactive power scheduling of each power plant in such a way as to minimize the total operating cost of the entire network [2]. In other words, the generator's real and reactive powers are allowed to vary within certain limits so as to meet a particular load demand with minimum fuel cost. This is called the Optimal Power Flow (OPF) or sometimes known as the Optimal Power Dispatch or Economic Dispatch (ED) problem [2], [3].

II. MODELLING OF OPTIMAL POWER FLOW (OPF) PROBLEM

In the solution of OPF, the main objective is to minimize total operating costs of the system. In OPF, when the load is light, the cheapest generators are always the ones chosen to run first. As the load increases, more and more expensive generators will then be brought in. Thus, the operating cost plays a very important role in the solution of OPF [5], [6].

In all practical cases, the cost of generator i can be represented as a cubic function of real power generation expressed in $/hr,

$$C_i = (\alpha_i + \beta P_i + \gamma_i P_i^2 + \xi_i P_i^3) * \text{fuel cost} \qquad (1)$$

Where Pi is the real power output of generator i, and α, β, γ and ξ are the cost coefficients.

Normally, the cost coefficients remain constant for a generator. The last term in the equation is the fuel cost, expressed in $/MBtu [7]. Another important characteristic of a generator is the incremental cost, also known as marginal cost. It is a measure of how costly it

will be to produce the next increment of power. The incremental cost can be obtained from the derivative of C_i of equation (1) with respect to P_i,

$$\frac{\partial C_i}{\partial P_i} = (\beta_i + 2\gamma_i P_i + 3\xi_i P_i^2) * \text{fuel cost} \qquad (2)$$

Which is expressed in $/MWhr.

The transmission losses become a major factor in a large interconnected network whereby power is being transmitted over long distances. A common function to represent total system real power losses in terms of the total real power output is the Kron's loss formula

$$P_L = \sum_{i=1}^{ng}\sum_{j=1}^{ng} P_i B_{ij} P_j + \sum_{i=1}^{ng} B_{oi} P_i + B_{oo} \qquad (3)$$

Where P_L is the total real power losses, and B_{ij} are the loss coefficients or B coefficients [8].

Optimal dispatch can be seen generally as a constrained optimization problem. When solving a constrained optimization problem, there are two general types of constraints, which are equality and inequality constraints. Equality constraints are constraints that always need to be enforced.

The constrained optimization problem can be solved using the Lagrange Multiplier method, and for simplicity, only the maximum and minimum real power limits are included as the inequality constraints.

The total operating cost of all generators in a system is given by,

$$C_i = \sum_{i=1}^{ng} C_i \qquad (4)$$

Where ng is the number of generator buses.

The total real power generation is then given by,

$$\sum_{i=1}^{ng} P_i = P_D + P_L \qquad (5)$$

Where $P_{i(min)} \leq P_i \leq P_{i(man)}$, P_D is the total real power demand, and P_L is the total system real power loss [9].

The Lagrange Multiplier can then be expressed as,

$$\ell = C_t + \lambda(P_D + P_L - \sum_{i=1}^{ng} P_i) + \sum_{i=1}^{ng} \mu_{i(max)} (P_i - P_{i(max)}) + \sum_{i=1}^{ng} \mu_{i(min)} (P_i - P_{i(min)}) \qquad (6)$$

Where the term second term is the equality constraint, while the last two terms are inequality constraints in equation (6) [10].

Note that both $\mu_{i(max)}$ and $\mu_{i(min)}$ are equal to zero if $P_{i(min)} \leq P_i \geq P_{i(max)}$, which means that the inequality constraints are inactive. The constraints will only be active when violated, which means $P_i > P_{i(max)}$ or $P_i < P_{i(min)}$. This is known as the Kuhn-Tucker necessary conditions of optimality, following the conditions below,

$$\frac{\partial \ell}{\partial P_i} = 0 \qquad (7)$$

$$\frac{\partial \ell}{\partial \lambda} = 0 \qquad (8)$$

$$\frac{\partial \ell}{\partial \mu_{i(min)}} = P_i - P_{i(min)} = 0 \qquad (9)$$

$$\frac{\partial \ell}{\partial \mu_{i(max)}} = P_i - P_{i(max)} = 0 \qquad (10)$$

The optimal solution can then be obtained by solving for the condition, $\frac{\partial \ell}{\partial P_i} = 0$ [11]-[13].

III. BACKGROUND OF ATC

The Available Transfer Capability (ATC) is a measure of the transfer capability remaining in the physical transmission network for further commercial activity over and above previously committed uses. Mathematically, ATC is defined as the Total Transfer Capability (TTC) less the Transmission Reliability Margin (TRM) and the Capacity Benefit Margin (CBM) [14]-[17].

$$\therefore \text{ATC} = \text{TTC} - \text{TRM} - \text{CBM} \qquad (11)$$

TTC is defined as the amount of electric power that can be transferred over the interconnected transmission network in a reliable manner while meeting all of a specific set of defined pre and post-contingency system conditions.

TRM is defined as that amount of transmission transfer capability necessary to ensure that the interconnected transmission network is secure under a reasonable range of uncertainties in system conditions.

CBM is defined as that amount of transmission transfer capability reserved by load serving entities to ensure access to generation from interconnected systems to meet generation reliability requirements.

References [18]-[20] provide additional analysis of errors associated with ATC.

IV. THE SPOT MARKET

Electricity in the National Electricity Market (NEM) can be either traded through retail or wholesale trading or even through contracts. Note that this paper

only emphasizes on the wholesale trading of the spot market. All wholesale electricity must be traded through the spot market; generators are paid for the electricity they sell to the pool while retailers and wholesale end-users pay for the electricity they use from the pool. It is a process whereby prices for electricity are set and then settled. This pool is the way which short-term operation of the power system is centrally.

In this spot market, generating companies can choose whether to commit their generators and make it available for dispatch. Once they have decided to commit, they must submit a bid to National Electricity Market Management Company Limited (NEMMCO) for the opportunity to run their generators. A bid is the "sell offer" submitted for a particular amount of electricity selling at a particular price. Generating companies can change their bids or submit re-bids according to a set of bidding rules. After receiving all the bids, NEMMCO will then selects the generators required to run and when to run at different times of the day, based on the most cost-efficient supply solution to meet specific demand. This ensures electricity is supplied at the lowest possible price. As mentioned above, the spot market allows instantaneous matching of supply against demand.

The spot price also varies within different states in Australia, where the electricity demand for each state varies, and is dependent on the population and the industrial and commercial mix.

A maximum spot price is also set under the Code, which is the maximum price, that generating companies can bid [21].

V. TEST CASES

The ideas are demonstrated by IEEE standard power systems: 26 and 118-buses, and results have been presented and analyzed.

A. 26-bus system

A 26-bus power system [2] is used for portfolio analysis in different operating scenarios. The generators' efficiencies and capabilities, types of generators owned, fuel costs, transmission losses and spot price variation are some of the factors that can affect generating companies' profit margins in a deregulated market environment. This section demonstrates that through proper analysis of these factors, generating companies can utilize solutions of OPF to maximize their profit margins through the wholesale spot market and is also extended to calculate ATC in this environment for different load conditions.

This section shows the power flow solutions of the 26-bus system, solved using the full Newton-Raphson (NR) algorithm, which includes the bus and generator data, total system costs, MW generation and losses, and results are given in Table I.

TABLE 1 : GENERATOR DATA OF THE 26-BUS SYSTEM (NR)

Gen no	Gen MW	Gen Mvar	Min MW	Max MW	Min Mvar	Max Mvar	Cost $/hr	IC $/MWhr
1	472	254	100	500	0	350	5109	13.61
2	171	78	50	200	40	250	2193	13.26
3	188	40	80	300	40	150	2141	11.89
4	148	80	50	150	40	80	2031	13.67
5	194	137	50	200	40	160	2564	13.61
26	101	27	50	120	15	50	1484	13.52

Total MW generation : 1276.5 MW

Total Mw load : 1263.0 MW

Total Mw losses : 13.52 MW

System cost : $16777.71 / hr

The power flow solutions of the 26-bus system solved using the OPF algorithm, and results are given in Table II.

Table –II
Generator data of the 26-bus System (OPF)

Gen no	Gen MW	Gen Mvar	Min MW	Max MW	Min Mvar	Max Mvar	Cost $/hr	IC $/MWhr
1	421	260	100	500	0	350	4429	12.90
2	170	65	50	200	40	250	2175	13.23
3	294	40	80	300	40	150	3496	13.79
4	130	80	50	150	40	80	1782	13.34
5	170	142	50	200	40	160	2236	13.22
26	92	28	50	120	15	50	1358	13.38

Total MW generation : 1277.0 MW

Total Mw load : 1263.0 MW

Total Mw losses : 14.03 MW

System cost : $15475.24 / hr

1) Case Scenarios

There are six generators in 26-bus system; these generators are distributed to two generating companies A and B. Company A owns generator 1, 4 and 26 while Company B owns 2, 3 and 5.

The generators are distributed in such a way that Company A owns the cheapest and most expensive generators while Company B owns those with moderate operating costs.

Simulations are done on with varying load conditions from 50% to 100% load

A. System at 100% Load

The total load demand is 1263 MW and maximum MW generation from all generators is 1520 MW. Assume all generators are committed for dispatch; NEMMCO will then selected the cheaper generators to run and to meet the demand. In this case, generator 26, which is most expensive, is not required. The maximum MW generation from the other five generators is 1400 MW, which is sufficient to meet the demand. The generators' bids are assumed to be 10% higher than the generators' costs and the spot price is determined by the highest generator's bid. Results are given in Table 3 and 4.

Profit ($/MWhr) = Spot Price($/MWhr) – Cost($/MWhr)

Profit ($/hr) = Profit ($/MWhr) x Gen MW

B. System at 90% Load

The total load demand is 1136.70 MW. In this case, generator 4 and 26 are not required. The maximum MW generation from the other four generators is 1250 MW, which is sufficient to meet the demand.

C. System at 75% Load

The total load demand is 947.25 MW. In this case, generator 4, 5 and 26 are not required. The maximum MW generation from the other three generators is 1050 MW, which is sufficient to meet the demand.

D. System at50% Load

The total load demand is 631.50 MW. In this case, generator 2, 4, 5 and 26 are not required. The maximum MW generation from the other two generators is 850 MW, which is sufficient to meet the demand.

Both Company A and B committed all their generators for dispatch. Company A had a much higher profit margin than Company B regardless of the load conditions. Table III below shows the comparison between both companies' profits and the spot price variation as load changes.

TABLE-III.
COMPARISION OF PROFITS OF COMPANY A AND B

MW Load %	Spot Price ($/MWhr)	Profits ($/hr)	
		Company A	Company B
50	12.617	884.09	248.9
75	14.08	1547.65	861.5
90	14.487	1748.5	1296.87
100	15.05	2240.3	1607

Results showed that profits are positively-related to the spot price and the load demand. In other words, profits increase as the spot price increases with the load demand. This is because the spot price is determined from the highest generator's bid, and expensive generators are required when the load demand is high, which will set a high spot price.

It is also realized that cheaper generators will have higher profit margins regardless of the spot prices. Therefore, it is advantageous for companies to own a greater number of cheap generators along with a few expensive ones. Those expensive generators can be used as backup units for emergencies and perhaps also used to set high spot prices which are beneficial to the cheaper generators.

From these results obtained, it can be concluded that types of generators owned by companies and that spot price variation can greatly affect their overall revenue.

ATC is calculated from all generator buses to other buses by using linear methods, and variation of ATC is given Tables IV to IX for different load conditions.

.

TABLE-IV. VARIATION OF ATC FROM BUS 1 TO OTHER BUSES

Transfer Buses	NR Method		100% Load		90% Load		75% Load		50% Load	
	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line
1-3	200	17-18	215	17-18	272	16-17	314	16-17	308	16-17
1-5	238	6-18	195	6-18	180	6-18	72	6-18	162	6-18
1-9	197	2-8	195	7-9	153	2-8	154	16-20	232	9-10
1-12	230	17-18	228	2-8	146	2-8	108	12-14	215	12-14
1-15	141	17-18	152	17-18	138	13-15	147	13-15	217	16-17
1-16	80	17-18	86	17-18	101	15-16	99	15-16	117	16-17
1-17	33	17-18	36	17-18	46	17-18	55	1-17	56	16-17
1-18	235	1-18	208	1-18	223	1-18	197	1-18	329	1-18
1-19	201	6-19	203	6-19	200	16-20	94	6-18	212	6-18
1-24	59	22-24	55	22-24	65	22-24	62	22-24	80	22-24

TABLE-V : VARIATION OF ATC FROM BUS 2 TO OTHER BUSES

Transfer Buses	NR Method		100% Load		90% Load		75% Load		50% Load	
	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line
2-1	666	1-2	672	1-2	691	1-2	670	1-2	723	1-2
2-3	251	17-18	270	17-18	333	16-17	350	2-8	378	16-17
2-5	266	6-18	218	6-18	201	6-18	81	6-18	181	6-18
2-9	193	7-9	194	7-9	139	2-8	112	2-8	232	9-10
2-12	224	2-8	208	2-8	134	2-8	105	12-14	209	12-14
2-15	156	12-15	175	12-15	135	13-15	144	13-15	230	13-15
2-16	86	17-18	90	15-16	98	15-16	96	15-16	123	15-16
2-17	34	17-18	37	17-18	47	17-18	52	16-17	54	16-17
2-18	259	1-18	229	1-18	246	1-18	217	1-18	363	16-17
2-19	204	6-19	205	6-19	189	2-8	109	6-18	246	6-18
2-24	59	22-24	54	22-24	65	22-24	62	22-24	79	22-24

TABLE-VI : VARIATION OF ATC FROM BUS 3 TO OTHER BUSES

Transfer Buses	NR Method		100% Load		90% Load		75% Load		50% Load	
	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line
3-1	225	3-13	182	3-13	104	3-13	94	12-14	188	12-14
3-5	206	3-13	166	3-13	95	3-13	76	12-14	152	12-15
3-9	193	3-13	157	3-13	86	12-14	62	12-14	124	12-14
3-12	183	12-14	150	3-13	69	12-14	50	12-14	100	12-14
3-15	139	13-15	124	13-15	72	3-13	76	3-13	158	3-13
3-16	84	15-16	78	15-16	77	3-13	80	3-13	106	15-16
3-17	40	17-18	42	17-18	52	16-17	44	16-17	45	16-17
3-18	185	16-17	167	16-17	99	3-13	89	12-14	155	16-17
3-19	195	6-7	160	3-13	91	3-13	69	12-14	138	12-14
3-24	57	22-24	53	22-24	63	22-24	60	22-24	78	22-24

TABLE – VII : VARIATION OF ATC FROM BUS 4 TO OTHER BUSES

Transfer Buses	NR Method		100% Load		90% Load		75% Load		50% Load	
	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line
4-1	216	4-12	218	4-12	386	4-12	279	6-7	342	16-17
4-3	194	4-12	195	4-12	346	4-12	357	4-12	359	12-15
4-5	196	6-7	161	6-7	205	6-7	98	6-7	144	6-7
4-9	178	10-12	162	10-12	219	7-9	218	9-10	218	9-10
4-12	162	4-12	163	4-12	289	4-12	299	4-12	313	4-12
4-15	118	12-15	133	12-15	178	13-15	190	13-15	190	12-15
4-16	92	15-16	85	15-16	92	15-16	91	15-16	116	15-16
4-17	38	17-18	40	17-18	52	17-18	47	16-17	48	16-17
4-18	214	4-12	215	4-12	240	16-17	178	6-7	206	16-17
4-19	177	10-12	162	10-12	232	6-19	149	6-7	218	6-7
4-24	56	22-24	52	22-24	62	22-24	59	22-24	76	22-24

TABLE-VIII : VARIATION OF ATC FROM BUS 5 TO OTHER BUSES

Transfer Buses	NR Method		100% Load		90% Load		75% Load		50% Load	
	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line
5-1	106	5-6	100	5-6	110	5-6	288	5-6	275	5-6
5-3	106	5-6	100	5-6	110	5-6	288	5-6	275	5-6
5-6	106	5-6	100	5-6	110	5-6	288	5-6	275	5-6
5-9	106	5-6	100	5-6	110	5-6	232	9-10	232	9-10
5-12	106	5-6	100	5-6	110	5-6	148	12-14	275	5-6
5-15	106	5-6	100	5-6	110	5-6	165	13-15	219	16-17
5-16	85	17-18	92	17-18	104	16-17	103	15-16	118	16-17
5-17	34	17-18	37	17-18	47	17-18	55	16-17	56	16-17
5-18	106	5-6	100	5-6	110	5-6	288	5-6	275	5-6
5-19	106	5-6	100	5-6	110	5-6	237	6-19	272	6-19
5-24	63	22-24	58	22-24	69	22-24	66	22-24	85	22-24

TABLE-IX : VARIATION OF ATC FROM BUS 26 TO OTHER BUSES

Transfer Buses	NR Method		100% Load		90% Load		75% Load		50% Load	
	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line	ATC (MW)	Limiting Line
26-1	82	11-26	185	11-26	181	11-26	169	11-26	175	11-26
26-3	81	11-26	183	11-26	180	11-26	168	11-26	173	11-26
26-5	76	11-26	170	11-26	168	11-26	156	11-26	162	11-26
26-6	76	11-26	170	11-26	168	11-26	156	11-26	162	11-26
26-9	80	11-26	178	11-26	175	11-26	163	11-26	169	11-26
26-12	80	11-26	179	11-26	176	11-26	128	11-26	170	11-26
26-15	81	11-26	166	12-15	147	13-15	156	13-15	171	11-26
26-16	80	11-26	94	17-18	102	15-16	100	15-16	123	16-17
26-17	34	17-18	37	17-18	47	17-18	53	16-17	55	16-17
26-18	80	11-26	179	11-26	176	11-26	164	11-26	170	11-26
26-19	76	11-26	171	11-26	168	11-26	156	11-26	162	11-26
26-24	61	22-24	56	22-24	67	22-24	64	22-24	83	22-24

.ATC numbers are varying with variation of load. This variation of ATC depends upon Generators in operation, Load variation, bus voltages, Variation of line flows and line limits.

B. IEEE 118-bus system

For the case study, a large sized system, IEEE 118-bus system is composed of 54 generators with a total installed capacity of 8190MW and the system demand is 3668MW. The system is divided into four areas as follows. For this analysis each area is considered as one generating company. The details of areas and dispatch of generators for different cases are given in Tables X and XII.

TABLE - X : DETAILS OF THE AREAS OF IEEE 118-BUS SYSTEM

Area	Buses	Company
1	1 – 38, 113 – 115 & 117	A
2	39 - 68	B
3	69 – 102, 116 & 118	C
4	103 - 112	D

This analysis is discussed under different case studies as follows

Case 1: All the generators are committed to dispatch with 100% load.

Case 2: Some of Generators with more expensive are shut down with 100% load.

Case 3: Some more Generators with more expensive are shut down with 75% load.

Case 4: Some more Generators with more expensive are shut down with 50% load

The details of generators are given in Table 19 below.

Comparisons of profits of companies A, B, C and D are given in Table XI.

TABLE - XI : Comparision of Profits of Company A , B, C and D

	Spot Price $/MWhr	Profit $/MWhr			
		Company A	Company B	Company C	Company D
Case 1	13.06	8334.77	9516.54	8780.16	1856.27
Case 2	9.35	3465.47	3989.82	4554.97	1213.00
Case 3	8.43	2014.43	2686.84	2860.80	745.16
Case 4	11.29	3185.91	3806.74	2301.03	888.62

In IEEE 118-bus test system also, results showed that profits are positively-related to the spot price and the load demand. But in the case of case 4, slack bus i.e. bus 69 highest generation cost, which increases the spot price and profits even though system load at 50%. By changing the slack bus from 69 to any other bus low cost of generation will reduce spot price and profits.

ATC is calculated between the areas by using linear methods, and variations of ATC numbers are given Table XIII. Thermal limits used for calculating ATC of all the lines is taken as 500MVA. The details of ATC between areas for different cases are given in Table XIII.

.

TABLE – XII : DISPATCH OF GENERATORS FOR DIFFERENT CASES

Area	Generators	Generators committed to dispatch			
		Case1	Case 2	Case 3	Case 4
1	1,4,6,8,10,12,15,18,19,24,25,26,27, 31,32,34,36,113	All	1,4,6,12,15,18,19, 26,34,36,113	12,15,18,19,26,34,36,113	12,15,18,19,34,113
2	40,42,46,49,54,55,56,59,61,62,65,66	All	40,42,49,54,55,56,59,62,66	40,42,54,55,56,59,62,66	40,54,55,56,59,62
3	69,70,72,73,74,76,77,80,85,87,89,90, 91,92,99,100,116	All	69,76,77,87,89,91,92,99,116	69,77,87,89,91,99	69,89,91
4	103,104,105,107, 110,111,112	All	110,111,112	110,111,112	110,112

TABLE – XIII : Variation of ATC between the areas for different cases of IEEE 118-bus System

Transfer Areas	Case 1		Case 2		Case 3		Case 4	
	ATC (MW)	Limiting Area/Line	ATC (MW)	Limiting Area/Line	ATC (MW)	Limiting Area/Line	ATC (MW)	Limiting Area/Line
1-2	706.87	Area 2	67.46	Area 1	236.40	Area 1	49.27	Area 1
1-3	127.24	Area 3	67.46	Area 1	236.40	Area 1	49.27	Area 1
1-4	101.59	Area 4	67.46	Area 1	110.87	Area 4	49.27	Area 1
2-1	454.57	Area 1	232.45	Area 2	288.61	Area 1	133.71	Area 2
2-3	127.24	Area 3	232.45	Area 2	291.71	Area 2	133.71	Area 2
2-4	101.59	Area 4	175.00	Area 4	110.87	Area 4	72.78	Area 4
3-1	454.57	Area 1	542.54	Area 1	288.61	Area 1	225.73	Area 1
3-2	706.87	Area 2	704.52	65-68	625.29	Area 2	261.44	Area 3
3-4	101.59	Area 4	175	Area 4	110.87	Area 4	72.78	Area 4
4-1	247.41	Area 4	0.00	Area 4	64.13	Area 4	27.22	Area 4
4-2	247.41	Area 4	0.00	Area 4	64.13	Area 4	27.22	Area 4
4-3	127.24	Area 3	0.00	Area 4	64.13	Area 4	27.22	Area 4

VI. CONCLUSION

In this paper, on IEEE 26-bus system and IEEE 118-bus system have been demonstrated for the purpose of calculating OPF and ATC calculations. This paper demonstrated that the proper scheduling of generators by using OPF minimized the total system losses and therefore increased generators efficiencies. It shows that the OPF algorithm had solved the case more cost-efficiently. Therefore increases the revenues of company in deregulated power system. It is also realized that cheaper generators will have higher profit margins regardless of the spot prices. Therefore, it is advantageous for companies to own a greater number of cheap generators along with a few expensive ones. Those expensive generators can be used as backup units for emergencies and perhaps also used to set high spot prices which are beneficial to the cheaper generators. From these results, it can be concluded that types of generators owned by companies and that spot price variation can greatly affect their overall revenue. The concept of Available Transfer Capability requires the determination of what is available from a particular condition. If the exact conditions were known in advance and a specific transaction was in question, the burden would be significantly less than that encountered in the attempt to predict what will be available at a future time. The results are certainly useful in an online environment of deregulated power system to perform the transactions between buyer and seller.

ACKNOWLEDGMENT

The authors are thankful to the management of CVR College of Engineering, Ibrahimpatnam, R.R District and JNT University, Hyderabad, for providing facilities and to publish this work.

REFERENCES

[1] R. H. Miller and J. H. Malinowski, “Economic Operation of Power Systems,” Power System Operation, 3rd Edition, McGraw-Hill, New York, 1970, pp. 63-82.

[2] H. Saadat, Power System Analysis, McGraw-Hill, Boston, 1999.

[3] H. H. Happ, “Optimal Power Dispatch,” IEEE Transactions on Power Apparatus and Systems, Vol. PAS-93, No. 3, May/June 1974, pp. 820-830.

[4] B. M. Weedy and B. J. Cory, “Basic Power-System Economics and Management,” Electric Power Systems, 4th Edition, John Wiley & Sons Ltd, New York, 1999, pp. 493-520.

[5] J. A. Momoh, R. J. Koessler and M. S. Bond, “Challenges to Optimal Power Flow,” IEEE Transactions on Power Systems, Vol. 12, No. 1, February 1997, pp. 444-455.

[6] A. M. Sasson and H. M. Merrill, “Some Applications of Optimization Techniques to Power Systems Problems,” Proceedings of the IEEE, Vol. 62, No. 7, July 1974, pp. 959-972.

[7] K. Y. Lee, Y. M. Park, and J. L. Ortiz, “Fuel-cost Minimization for both Real and Reactive-power Dispatches,” IEE Proceedings, Vol. 131, Pt. C, No. 3, May 1984, pp. 85-93.

[8] J.J. Grainger and W. D. Stevenson, Power System Analysis, McGraw-Hill, New York, 1994.

[9] "Common Format for Exchange of Solved Load Flow Data," IEEE Transactions on Power Apparatus and Systems, Vol. PAS-92, 1973, pp. 1916-25.

[10] J. A. Momoh, M. E. El-Hawary and R. Adapa, “A Review of Selected Optimal Power Flow Literature to 1993 Part I: Nonlinear and Quadratic Programming Approaches,” IEEE Transaction on Power Systems, Vol. 14, No. 1, February 1999, pp. 96-104.

[11] D. I. Sun, B. Ashley, B. Brewer, A. Hughes and W. F. Tinney, “Optimal Power Flow By Newton Approach,” IEEE Transaction on Power Apparatus and Systems, Vol. PAS-103, No. 10, October 1984, pp. 2864-2880.

[12] H. H. Happ, “Optimal Power Dispatch – A Comprehensive Survey,” IEEE Transactions on May/June 1977, pp. 841-854.

[13] H. W. Dommel and W. F. Tinney, “Optimal Power Flow Solutions,” IEEE Transactions on Power Apparatus and Systems, Vol. PAS-87, October 1968, pp. 1866-1876.

[14] NERC Operating Manual (latest edition), available at http:www.nerc.com

[15] Transmission Transfer Capability Task Force, Transmission Transfer Capability, North American Electric Reliability Council, Princeton, New Jersey, May 1995.

[16] Transmission Transfer Capability Task Force, Available Transfer Capability Definitions and Determination, North American Electric Reliability Council, Princeton, New Jersey, June 1996.

[17] Mojgan Hojabri , Hashim Hizam , Norman Mariun , Senan Mahmood Abdullah, Available Transfer Capability (ATC): A Review, International Review of Modelling and Simulations, Vol.2 (Issue 4); 351-355, AUGUST 2009.

[18] P. W. Sauer, Technical Challenges of Computing Available Transfer Capability (ATC) in Electric Power Systems, Proceedings of the Thirtieth Annual (1997) Hawaii International Conference on System Sciences, Vol. V, pp. 589-593, Maui, Hawaii, January 7- 10, 1997.

[19] P. W. Sauer, Alternatives for calculating Transmission Reliability

[20] Margin (TRM) in Available Transfer Capability (ATC), Proceedings of the Thirty-First Annual (1998) Hawaii International Conference on System Sciences, Vol. III, Kona, Hawaii, January 6-9, 1998, p. 89.

[21] S. Grijalva and P. W. Sauer, Reactive Power Considerations in Linear ATC Computation, Proceedings of the Thirty-Second Annual (1999) Hawaii International Conference on System Sciences, pp. 1- 11, Maui, Hawaii, January 5-8, 1999.

[22] An Introduction to Australia's National Electricity Market, National Electricity Market Management Company Limited, March 2001.

❑❑❑

Differential Diagnosis of Dementia Using Slantlet Transform

M.M.Patil[1] & A.R.Yardi[2]

Department of Electronics and Telecommunication,College of Engg.Pandharpur , Dist.Solapur, Maharashtra, India.
Department of Electronics, Walchand College of Engg, Sangli, Maharashtra, India.
E-mail: mns.patil@gmail.com, aryardi@yahoo.com

***Abstract* -** The paper proposes use of a new approach for automated multiclass diagnosis, based on classification of magnetic resonance images (MRI) of human brain. Wavelet transform based methods are a well-known tool for extracting frequency space information from non-stationary signals. We employ an improved version of orthogonal discrete wavelet transform (DWT) for feature extraction, called Slantlet transform, which can especially be useful to provide improved time localization with simultaneous achievement of shorter supports for the filters. For each two-dimensional MR image, we have computed its intensity histogram and Slantlet transform has been applied on this histogram signal. Then a feature vector, for each image, is created by selecting the six absolute largest Slantlet outputs. The features hence derived are used to train a neural network based four class classifier, which can automatically infer whether the MR image belongs to a normal brain or to a person suffering from Alzheimer's disease or Mild Alzheimer's disease or Huntington's Disease. An excellent classification ratio of 100% is achieved for a set of benchmark MR brain images, which was significantly better than the results reported in a very recent research work employing wavelet transform and neural networks. Used Matlab 7 in both stages, feature extraction from 2D MRI as well as neural network part.

Keywords- *Magnetic resonance imaging (MRI); Feature extraction; Classification; Slantlet transform (ST); Supervised neural network*

I. INTRODUCTION

A. Differential Diagnosis of Dementia

Dementia refers to the loss of memory and other cognitive skills due to changes in the brain caused by disease or trauma. The changes can affect thinking, memory and reasoning, and may occur gradually or quickly. An estimate tells 24 million people worldwide suffer from dementia—and the numbers are growing. To make matters worse, many people don't understand the difference between Alzheimer's and other forms of dementia, causing many cases to go undiagnosed and untreated [1]. Hence there is an urgent need to understand the disease, to develop prophylactic strategies which can distinguish the different types of Dementia, so that the proper step can be taken to manage the disease. Magnetic resonance imaging (MRI) is a powerful and flexible medical imaging modality. Among many other capabilities, it can produce high-resolution images with good contrast of the different biological soft tissue types. As a non-invasive technique, it is widely used in the clinical and research environments for imaging both anatomy and function [2]. Hence we planned to use MR images for the differential diagnosis of dementia.

B. Slantlet transform

As MRA (Multi Resolutin Analysis) is designed to give good time resolution and poor frequency resolution at high frequencies and good frequency resolution and poor time resolution at low frequencies. Good for signal having high frequency components for short durations and low frequency components for long duration. e.g. images and video frames , planned to find proper wavelet transform which could track the MRI irregularities occurring in various Dementia types.

The discrete wavelet transform is usually carried out by filter bank iteration; however, for a fixed number of zero moments, this does not yield a discrete-time basis that is optimal with respect to time-localization. W. Selesnick [3] discussed the implementation and properties of an orthogonal DWT, with two zero moments and with improved time-localization. The basis is not based on filter bank iteration; instead different filters are used for each scale. For coarse scales, the support of the discrete-time basis functions approaches 2/3 that of the corresponding functions obtained by filter bank iteration. This basis, a special case of a class of bases described by Alpert, retains the octave-band characteristic and is piecewise linear (but

discontinuous). The Slantlet transform (ST) thus been recently proposed as an improvement over the classical DWT, which can provide better time localization. They are inspired by an equivalent form of DWT, where the filter bank structure is implemented in a parallel form, employing different filters for each scale. ST employs a special case of a class of bases. ST can be implemented employing filters of shorter supports and, yet maintaining the desirable characteristics like orthogonality and an octave-band characteristic, with two zero moments. Fig.1 shows two scale orthogonal DWT iterated filter bank with two zero moments.

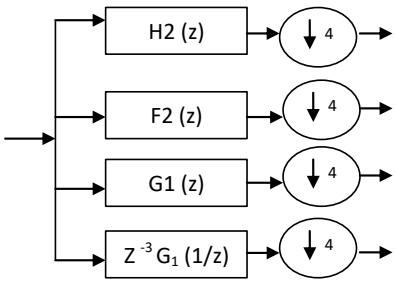


Fig. 1 : Two scale orthogonal DWT iterated filter bank (Slantlet filter bank)

Here for l scales let gi(n), fi(n) and hi(n) be the filters employed in scale i to analyze the signal. Each of these filters has an exact support of 2i+1. For l scales, ST filter bank employs l number of channel pairs, i.e. a total of 2l number of channels. Hence, the low pass filter $H_l(n)$ is paired with its adjacent filter $F_l(n)$ and each filter is followed by a downsampling by 2^l. Each of the other (l-1) channel pairs constitutes of a Gi(n) filter and its shifted time reversed version (i = 1, 2, . . ., l- 1), followed by a downsampling by 2^{i+1}. As the filters gi(n), fi(n) and hi(n) are implemented in piecewise linear forms, they can be represented by following expressions:

$$g_i(n) = \begin{cases} a_{0,0} + a_{0,1}n, & \text{for } n = 0, \ldots, 2^i - 1 \\ a_{1,0} + a_{1,1}n, & \text{for } n = 2^i, \ldots, 2^{i+1} - 1 \end{cases}$$

$$h_i(n) = \begin{cases} b_{0,0} + b_{0,1}n, & \text{for } n = 0, \ldots, 2^i - 1 \\ b_{1,0} + b_{1,1}n, & \text{for } n = 2^i, \ldots, 2^{i+1} - 1 \end{cases}$$

$$f_i(n) = \begin{cases} c_{0,0} + c_{0,1}n, & \text{for } n = 0, \ldots, 2^i - 1 \\ c_{1,0} + c_{1,1}n, & \text{for } n = 2^i, \ldots, 2^{i+1} - 1 \end{cases}$$

Hence the procedure of the filter design essentially reduces to the determination of the 12 parameters as above - as, bs and cs, for each ith scale. This determination is based on the formulation of several constraints, which satisfy orthogonality and two vanishing moments. [4]

II. METHOD

A. *Feature Extraction*

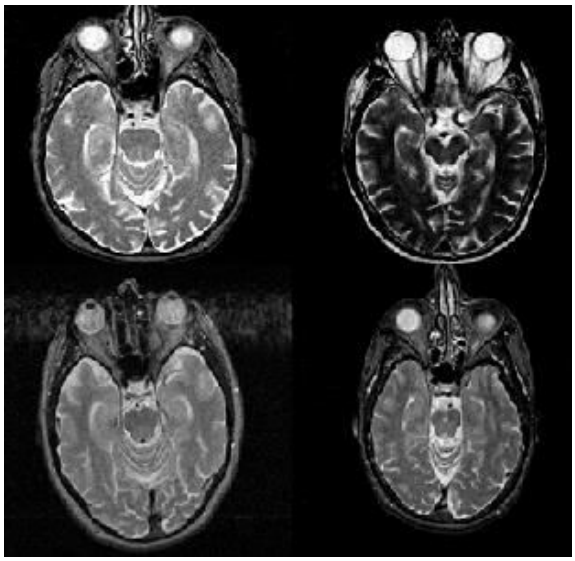


Fig. 2 : MR image slice for AD (top left), Mild AD (top right), Huntington's Disease (bottom left), Normal(bottom right)

Fig.2 shows 2D MR images of AD, Mild AD, Huntington's disease Dementia and Normal brain from the Whole brain atlas database[5].

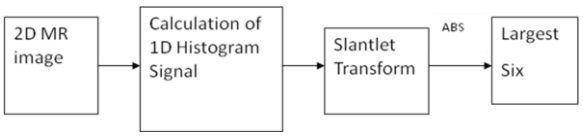


Fig. 3 : Feature extraction process

In the present research work, for each two-dimensional MR brain image, we have first created the one dimensional histogram of the intensities and then we have utilized ST to extract features from these histogram signals, each treated as a finite length data signal. For each such signal (derived from each corresponding image), we have extracted six absolute largest Slantlet transform outputs and then we have implemented a supervised neural network based classifier. Numbers of experiments were performed in an attempt to carry out differential diagnosis of Dementia into various subtypes. The histogram signal obtained for each disease type is shown in Fig.4. The Slantlet transformation to give a list of Slantlet coefficients as plotted in Fig.5. Feature reduction: Slantlet outputs first are converted to absolute values and then selected the largest six Slantlet coefficients as features. Process is repeated for every image in the experiment.

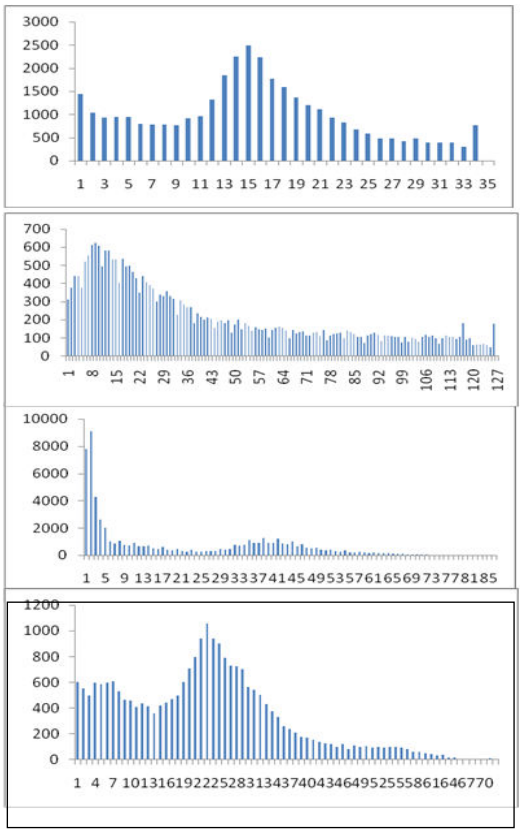
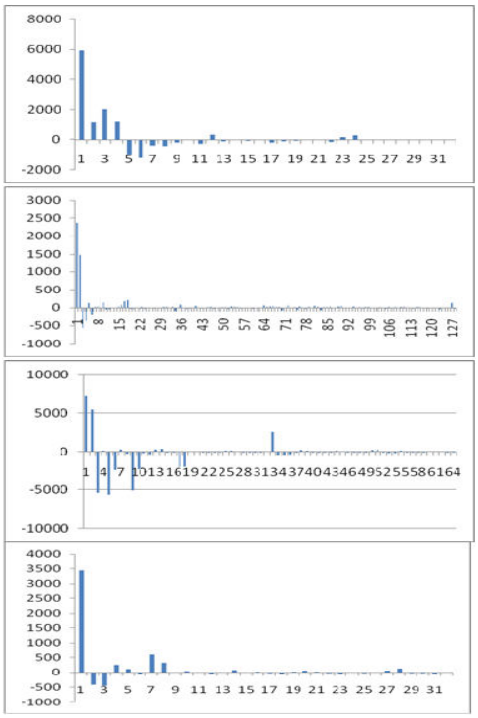


Fig. 4 : Intensity histogram signals for AD, Mild AD, HD and Normal, respectively from top (X –voxel position,Y-Intensity)



Slantlet outputs calculated from Intensity histogram signals for AD, Mild AD, HD and Normal, respectively from top to bottom(X –voxel position,Y-Intensity)

B. ANN Design

Divided the features set into training (81) and testing (41) sets corresponding to four disease classes- AD, Mild AD, Huntington's Disease and Normals. Decided to use Feed forward neural network. By giving training inputs, training experiment is carried out several times by varying various ANN parameters like activation functions, number of neurons at hidden layer, learning rate. Final configuration which gave 100% accuracy (Tab.5.8.2) was with 12 hidden neurons and tansig-logsig-purelin combination of activations (Fig.6 and Fig.7), took 351 epoches to reduce mse to 1.4E-11 with initial learning rate of 0.001.

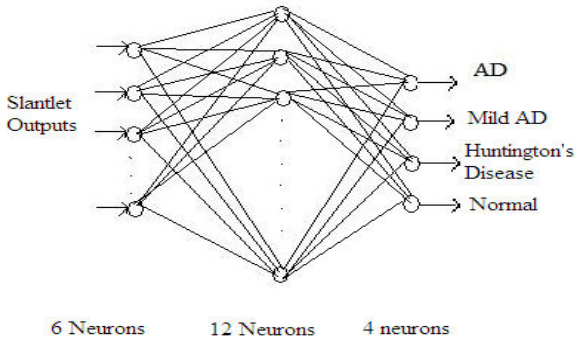


Fig. 6 : ANN architecture used

Finally achieved 100% success for four classes experiment.Performance is evaluated every time by applying test set and comparison is made with other results.

III. RESULTS

FFNN is now tested by applying various test cases and found to give accurate results. The performance of the proposed algorithm was evaluated by applying several test cases and computing the percentages of Sensitivity (SE), Positive Predictivity and Accuracy (AC). The following formulas:

$$SE = \frac{TP \times 100}{(TP+FN)}$$

$$PP = \frac{TPx100}{TP+FP}$$

$$AC = \frac{(TP+TN) \times 100}{(TN+TP+FN+FP)}$$

where, (class: AD, Mild AD, Huntington's Disease and Normals)

FP: Predicts non-class as of class.

TP: Predicts class as class.

FN: Predicts class as non-nonclass.

TN: Predicts non-class as non-class.

The calculated SE, PP and AC were all 100% as can be verified from testing result presented in Tab.I.

IV. CONCLUSION AND DISCUSSION

- Proposed the development of an automated brain MRI diagnostic system, which can classify whether the MR image belongs to a normal brain or to a person suffering from Alzheimer's disease or Mild Alzheimer's disease or Huntington's Disease.
- The system implemented a two-stage algorithm. In stage 1, for each image, its intensity histogram is computed and then Slantlet transform is employed to extract six features from this intensity histogram. In stage 2, a supervised neural network based classifier is developed, on the basis of the extracted features, to perform classification.
- ST has been very recently developed as an improvement over DWT, which can provide shorter supports for component filters and hence facilitates better time localization.
- The proposed system could provide an excellent 4 class classification accuracy of 100% by utilizing as low as six features only for the classifier input. This showed significant improvement compared to some of the results reported very recently, based on benchmark images acquired from the same database [6],[7],[8],[9].

.

TABLE-1 : **TEST RESULT OF FOUR CLASS CLASSIFIER**

Test I/P	Slantlet 1	Slantlet 2	Slantlet 3	Slantlet 4	Slantlet 5	Slantlet 6	Target				Actual			
AD1	0.54909	0.51463	0.17828	0.16091	0.064065	0.032945	1	0	0	0	1	1.7E-06	-8.5E-06	1.03E-06
AD2	0.51138	0.50352	0.15415	0.065164	0.058287	0.041822	1	0	0	0	1	-2.4E-06	-4.5E-06	-4.6E-06
AD3	0.53424	0.31001	0.0629	0.054944	0.046706	0.028294	1	0	0	0	1	-1.5E-06	1.09E-06	-3.8E-06
AD4	0.53565	0.28981	0.077525	0.077511	0.040421	0.038095	1	0	0	0	1	-1.2E-06	1.08E-06	-1.4E-06
AD5	0.57219	0.10444	0.099963	0.090247	0.053291	0.045103	1	0	0	0	0.99998	1.15E-05	5.15E-06	7.96E-07
AD6	0.58022	0.11444	0.11181	0.10894	0.08058	0.071627	1	0	0	0	0.99999	8.32E-06	2.72E-06	-1.7E-06
AD7	0.55729	0.26073	0.19624	0.12819	0.12378	0.096689	1	0	0	0	1	-1.3E-06	-1E-05	6.8E-06
AD8	0.54675	0.24531	0.19393	0.14044	0.12478	0.10306	1	0	0	0	1	-1.6E-06	-9.8E-06	6.59E-06
AD9	0.55554	0.23424	0.18196	0.1275	0.11058	0.10935	1	0	0	0	1	-2E-06	-6.9E-06	5.81E-06
AD10	0.44926	0.11686	0.094755	0.07918	0.056684	0.03278	1	0	0	0	1	1.38E-05	3.55E-06	-1.3E-05
AD11	0.44507	0.083102	0.081387	0.07614	0.054685	0.051254	1	0	0	0	0.99998	-5.2E-07	6.85E-06	9.85E-06
MldAD1	0.20559	0.13667	0.024306	0.018343	0.015053	0.009269	0	1	0	0	1.44E-06	1	7.62E-07	6.45E-07
MldAD2	0.19228	0.12001	0.045201	0.014502	0.010455	0.010135	0	1	0	0	-2.7E-05	1.0001	2.84E-06	-0.00012
MldAD3	0.19013	0.1108	0.038157	0.010934	0.010829	0.009645	0	1	0	0	-1.4E-05	1.0009	4.39E-06	-0.00087
MldAD4	0.22528	0.1737	0.04704	0.04701	0.036226	0.033491	0	1	0	0	2.92E-07	1	1.22E-06	-2.5E-06
MldAD5	0.22798	0.16404	0.062977	0.048563	0.045159	0.021249	0	1	0	0	4.72E-07	0.99999	1.98E-06	3.86E-06
MldAD6	0.24797	0.19002	0.067124	0.066294	0.05309	0.034489	0	1	0	0	-3.2E-07	1	2.49E-06	1E-08
MldAD7	0.24143	0.18244	0.1034	0.067731	0.061784	0.046486	0	1	0	0	-8.4E-07	1	1.53E-06	-2.9E-07
MldAD8	0.193	0.12654	0.068704	0.062966	0.031349	0.029139	0	1	0	0	-1.1E-05	1	-5.3E-07	-5.1E-06
MldAD9	0.17526	0.089244	0.050803	0.048182	0.03037	0.025237	0	1	0	0	-7.3E-05	0.99695	1.32E-06	0.003126
MldAD10	0.15654	0.075097	0.037809	0.037794	0.029552	0.023883	0	1	0	0	1.41E-05	1.0049	-1E-06	-0.00493
HD1	0.78851	0.7456	0.72975	0.63795	0.57074	0.31916	0	0	1	0	7.29E-07	1.06E-06	1	-8.6E-07
HD2	0.75515	0.75333	0.73946	0.62146	0.60814	0.33121	0	0	1	0	-2E-06	-1.4E-06	1	2.47E-06
HD3	0.73688	0.70846	0.67972	0.57921	0.55526	0.28558	0	0	1	0	-1.8E-06	-5.7E-07	1	2.04E-06
HD4	0.69084	0.6813	0.62849	0.5889	0.48388	0.28894	0	0	1	0	1.39E-06	2.5E-06	1	-4.1E-06
HD5	0.70279	0.68503	0.67153	0.61955	0.47936	0.29861	0	0	1	0	-4.2E-06	-1.9E-06	1	2.82E-06
HD6	0.72422	0.68675	0.66033	0.62275	0.5093	0.3096	0	0	1	0	-3.3E-06	-1.5E-06	1	2.41E-06
HD7	0.73893	0.60562	0.60377	0.53755	0.49231	0.25767	0	0	1	0	1.11E-06	1.64E-06	1	-8.5E-07
HD8	0.72939	0.55848	0.54685	0.54036	0.50712	0.23474	0	0	1	0	-6E-07	5.04E-07	1	7.35E-07
Nor1	0.34021	0.24467	0.042038	0.018889	0.018536	0.016345	0	0	0	1	1.04E-05	-0.0195	-2.6E-06	1.0195
Nor2	0.33678	0.20968	0.01226	0.00968	0.006864	0.006115	0	0	0	1	9.82E-05	-0.07299	-2.8E-06	1.0729
Nor3	0.33428	0.16864	0.015047	0.011689	0.010039	0.00738	0	0	0	1	0.000113	-0.00559	-5.4E-07	1.0055
Nor4	0.33202	0.061219	0.026616	0.023635	0.012855	0.010073	0	0	0	1	1.02E-06	-7.7E-06	-2.7E-07	1
Nor5	0.34376	0.029248	0.023707	0.021934	0.021738	0.011344	0	0	0	1	-4.1E-06	-7.2E-06	3.7E-07	1
Nor6	0.3425	0.031981	0.022234	0.022036	0.008741	0.008331	0	0	0	1	-3.2E-06	-8.8E-06	3.11E-07	1
Nor7	0.32969	0.069307	0.066783	0.042888	0.036983	0.01275	0	0	0	1	-5.2E-06	3.77E-06	7.5E-07	1
Nor8	0.31928	0.076024	0.064841	0.054466	0.037338	0.015436	0	0	0	1	2.02E-08	1.92E-09	7.8E-09	1
Nor9	0.30826	0.10357	0.097484	0.081161	0.052621	0.030299	0	0	0	1	2.8E-05	-0.00021	1.23E-07	1.0002
Nor10	0.3054	0.10253	0.093228	0.074605	0.0553	0.038065	0	0	0	1	9.09E-05	-0.00015	-7.8E-07	1.0001
Nor11	0.28884	0.093275	0.087626	0.079282	0.050668	0.027552	0	0	0	1	1.16E-05	-1.8E-05	-1.3E-06	1
Nor12	0.28753	0.09406	0.091718	0.080179	0.045313	0.025369	0	0	0	1	7.29E-06	-7.1E-05	-9.4E-07	1.0001

. **ACKNOWLEDGMENT**

Authors are thankful to the Mathworks for the Matlab software and the knowledgebase at www.mathworks.in/. Authors are also thankful to the Whole brain atlas database at http:// www.med.harvard.edu/ AANLIB/nav.htm., for the images.

REFERENCES

[1] http://www.helpguide.org/elder/alzheimers_disease_symptoms_stages.htm

[2] Helen Chiu Culture and Dementia From Diversity to Partnership Department of Psychiatry The Chinese University of Hong Kon Alziemers association, the year in alzheimer science 2008 Strides in Early Detection of Alzheimer's Disease W. Selesnick, The slantlet transform, IEEE Trans on Signal Processing, vol 47, no 5, pp 1304-1313, May 1999.

[3] W. Selesnick, The slantlet transform, IEEE Trans on Signal Processing, vol 47, no 5, pp 1304-1313, May 1999.

[4] Madhubanti Maitra, Amitava Chatterjee, A Slantlet transform based intelligent system for magnetic resonance brain image classification, Biomedical Signal Processing and Control 1 (2006) 299–306 doi:10.1016/j.bspc.2006.12.001

[5] http://www.med.harvard.edu/AANLIB/nav.htm.

[6] R.K. Begg, M. Palaniswami, B. Owen, Support vector machines for automated gait classification, IEEE Biomed. Eng. 52 (5) (2005) 28– 838.

[7] C.L. Dent, J.P. Spaeth, B.V. Jones, S.M. Schwartz, T.A. Glauser, B. Hullinan, J.M. Pearl, P. Khuri, C.D. Kurth, Brain magnetic resonance imaging abnormalities after the Norwood procedure using regional cerebral perfusion, J. Thorac. Cardiovasc. Surg. 130 (6) (2005) 1523–1530.

[8] R. Katz-Brull, A.R. Koudinov, H. Degani, Direct detection of brain acetylcholine synthesis by magnetic resonance spectroscopy, Brain Res. 1048 (1–2) (2005) 202–210.

[9] S. Chaplot, L.M. Patnaik, N.R. Jagannathan, Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network, Biomed. Signal Proc. Cont. 1 (2006) 86–92.

◈◈◈

Modeling of Novel Control Strategy used in Active Harmonic Filter for harmonic mitigation and reactive power compensation

Vivek Vora & Ramanujam Cadatur

Bombardier Transportation India Ltd, Vadodara, India

E-mail : vivek.vora@in.transport.bombardier.com, ramanujam.cadatur@in.transport.bombardier.com

Abstract - An Active harmonic filter is used to eliminate current harmonics caused by nonlinear loads. In this paper a 3-phase active harmonic filter based on hysteresis current control scheme is proposed. The proposed control technique is simulated using MATLAB. Simulation study and analysis shows that control technique is simple in implementation and fast in operation.

Keywords - Component; Active harmonic filter (AHF), nonlinear load (NLL), Voltage Source Inverter (VSI), PWM, Propotional Integral (PI), Total Harmonic Distortion (THD).

I. INTRODUCTION

Nowadays solid state power converters are widely used in application such as adjustable speed drive, uninterruptable power supply, Arc and Induction furnaces, and asynchronous AC-DC link in wind generation. These power converters behave as a non linear load to the supply mains. The nonlinear load injects current harmonics to mains and draw reactive component from mains. This injected current harmonics and reactive power component cause low power factor, low efficiency, and poor utilization of distribution system. Injected harmonics cause generation of voltage harmonics when passed through source impedance. Conventionally passive filter composed of L-C were employed to reduce harmonics and improve power factor. This passive filter have disadvantage of larger size, resonance and fixed compensation. However in some application were amplitude and harmonic content of distortion power vary randomly, this conventional solution becomes ineffective. In last couple of decades the concept of Active Harmonic Filter (AHF) has been introduced. There are various control approach for active filtering like, instantaneous power theory, notch filter, flux based controllers [1]-[3], [7], [14]. Most of these control algorithms are difficult to implement, slow in response as many transformation is included. This paper presents a simple algorithm to achieve control of AHF.

The control approach is simulated in MATLAB, realization of non linear load is obtain by three phase uncontrolled rectifier with Resistive loading. The algorithm makes the control simple, fast and stable in steady state and transient behavior.

II. BASIC CONFIGURATION

The basic configuration of active harmonic filter is as shown in fig. 1. AHF is composed of standard 3-phase voltage source Inverter Bridge with DC link capacitor to provide an effective current control [6], [12]. Nonlinear load composed of three phase uncontrolled rectifier with Resistor draws harmonic current from the source. Here I_S is the source current, I_L is load current and I_F is filter compensating current. The AHF is connected in parallel with nonlinear load. AHF calculates the harmonic current from load current. Source voltage synchronization is applied to derive compensating current [9], [10].

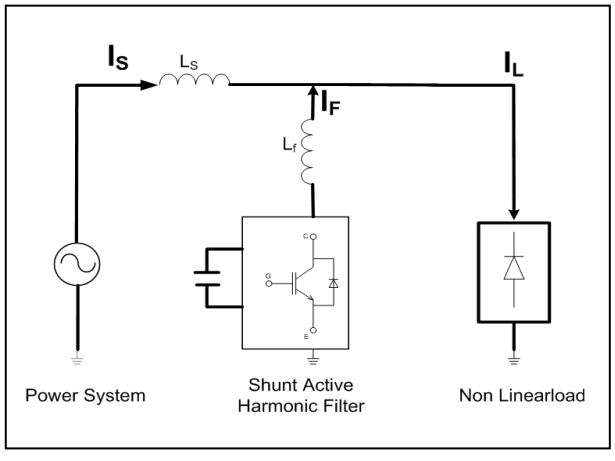


Fig. 1 : Basic Configuration of AHF.

Measurement of harmonic signal called total harmonic distortion is expressed mathematically as below.

$$THD = \frac{\sqrt{\sum_{n=2}^{\infty} I_h^{\,2}}}{I_1}$$

Where I_h is the harmonic current of rank h and I_1 is the fundamental current.

III. CONTROL ALGORITHM

A. Harmonic detection

The proposed control algorithm is simple in implementation and fast in operation compared to other control technique for AHF. Block diagram of control scheme is shown in fig. 2. Load current is sensed through current sensor and given as an analog input to the control algorithm. There are various techniques available for calculating maximum value of fundamental current I_m [2], [4], [8]. This control use sin multiplication method to derive I_m. Using the maximum value of fundamental current, the instantaneous fundamental current for active and reactive component is calculated. Harmonic current is calculated by subtracting fundamental currents from the load current.

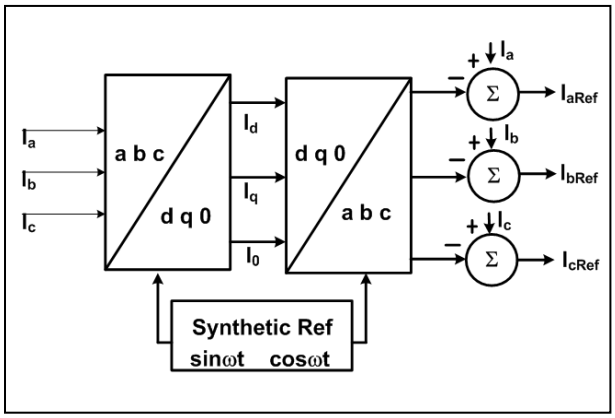


Fig. 2 : Control algorithm.

B. PWM generation for Voltage source inverter

IGBT based voltage source inverter (VSI) is operated in such a way that it will inject harmonic current of 180° out of phase to load harmonic, and draw active current to charge DC link and supply switching losses of IGBTs [6], [12].

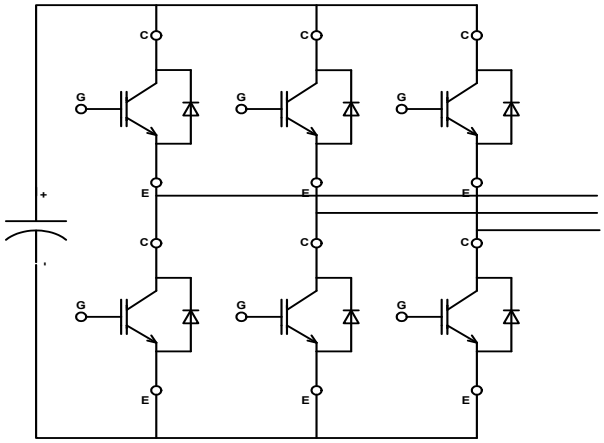


Fig. 3 : Voltage Source Inverter.

For controlling VSI, hysteresis current control method is applied so as output filter current follow reference current. PI controller ensures smooth control over wide range [10], [11], [15].

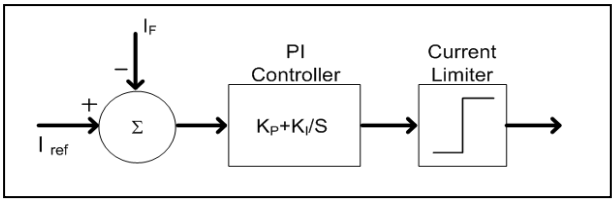


Fig. 4 : Current Control Loop using PI.

In order to meet the losses in IGBT switching and to stabilize the DC voltage a small active current is drawn from the supply, which is in phase with voltage. For stabilizing the DC voltage a voltage control loop with PI controller is implemented [12]. The Block diagram for the combined voltage control, and current control loop is shown in fig. 5.

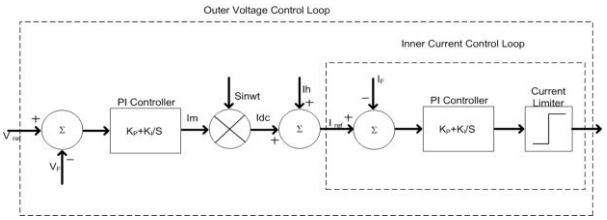


Fig. 5 : Combined Voltage-Current Control Loop.

IV. SYSTEM SIMULATION

Fig. 6. Shows block diagram of MATLAB simulation model. The values used in simulation are depicted in table.

Table 1 : Active filter parameters

Sr. No.	Simulation Parameters		
	Parameter	***value***	***unit***
1.	Supply Voltage $V_A=V_B=V_C$	230	V_{rms}
2.	Supply Frequency	50	H_Z
3.	Filter inductor	1	mH
4.	Filter Resistor	0.001	Ω
5.	Capacitor	10	mF
6.	Inductor	0.01	H
7.	Resistor	0.1	mΩ

In simulation it is assumed that DC Link capacitor is initially charged to a certain value in order to avoid high current at starting. For implementation it can be achieved via soft starting. Switching frequency of the converter is selected to 12.8 $KH_{Z,}$ which fulfill the criteria of realistic approach towards available IGBTs. Filter inductor ensure current to ramp up and down with reduced ripple level.

$$L = \frac{V_{dc} - V_m}{I_m \omega_0} \quad where\ \omega_0 = 2\pi f$$

$$C = \frac{I_m}{\%ripple \times V_{DC} \times \omega_0 \times 2}$$

.

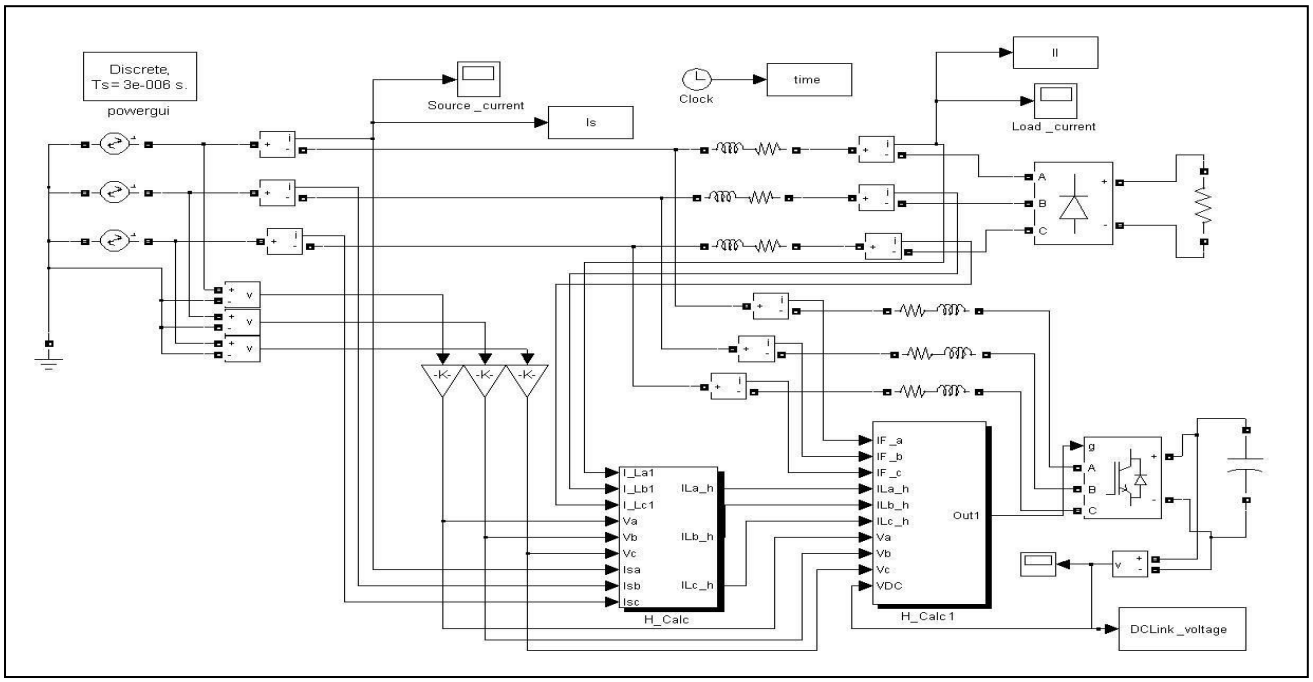


Fig. 6 : Simulation model

.

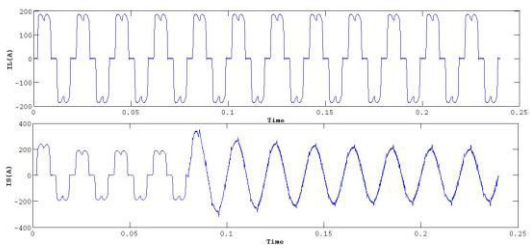


Fig. 7 (a) Load current and Source Current.

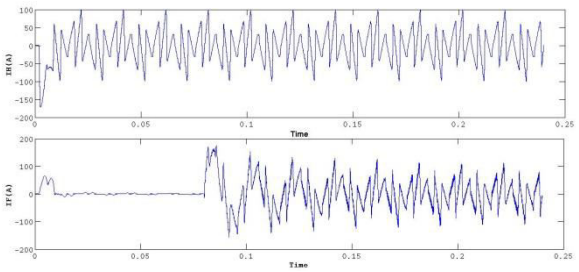


Fig. 7 (b) Calculated Harmonic Current and Filter Current

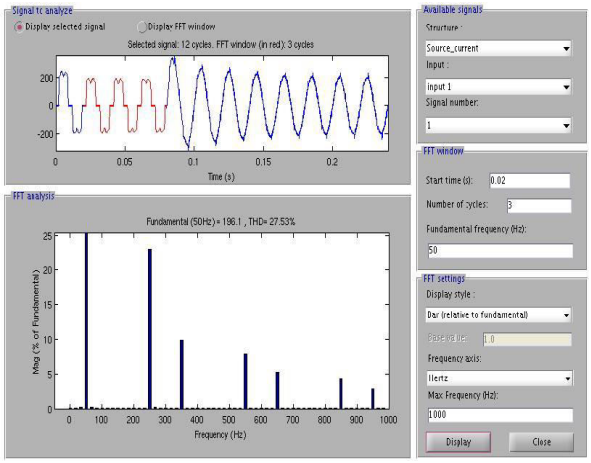


Fig. 8 (a) : THD Spectrum without AHF.

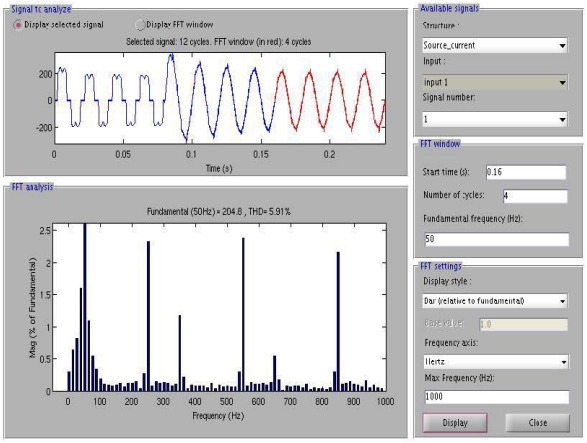


Fig. 8 (b) THD Spectrum with AHF

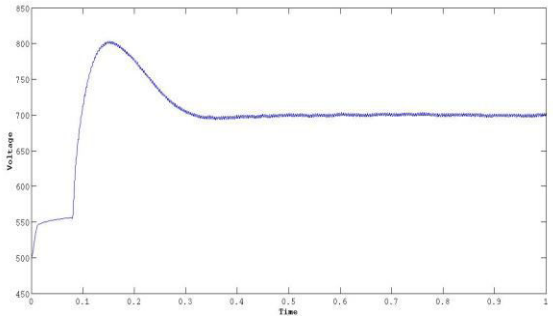


Fig. 9 : DC Link Voltage.

Comparative analysis of THD with and without AHF is depicted in table below. It can be seen that THD is reducing drastically from 27.53% to 5.91%.

TABLE : % HARMONIC CONTENT OF SUPPLY CURRENT

Harmonic Order	*Without AHF (%)*	*With AHF (%)*
5^{th}	22.99	2.32
7^{th}	9.92	1.18
11^{th}	7.94	2.39
13^{th}	5.28	0.54
17^{th}	4.31	2.16
19^{th}	2.84	0.05
23^{th}	2.45	1.79
25^{th}	1.53	0.38
THD	**27.53**	**5.91**

V. CONCLUSION

The proposed control scheme for active harmonic filter has high performance characteristic. Simulation results shows performance of AHF is quite satisfactory with this control, it reduces the THD value to less than 6%, and maintain sinusoidal source current. The control is efficient and simple for implementation.

REFERENCES

[1] J. Jacobs, D. O. Detjen, R. W. De Doncker, "An overview of methods to determine the harmonics in three-phase systems", IEEE Young Researchers Symposium in Electrical Power Engineering Distributed Generation, 2002.

[2] N. Mendalek. Al-Haddad, F. Fnaiech, L.A. Dessaint, "Sliding mode control of 3-phase 3-wire shunt active filter in the dq frame". Proc. Of Electrical and Computer Engineering 2001 Vol. 2, 2001, pp.765-769.

[3] H. Rudnick, J. Dixon, L. Moran, "Active power filters as a solution to power quality problems in distribution networks", Toulouse EPE 2003.

[4] Chandra, A., Singh, B., Singh, B. N., Al-Haddad, K., "An Improved Control Algorithm of Shunt Active Filter for Voltage Regulation, Harmonic Elimination, Power-factor Correction and Balancing of Nonlinear Loads",IEEE Transaction on Power Electronics, Vol. 15, No. 3, pp. 495-507, 2000.

[5] Salo M., Tussa H., "A Novel Open-Loop Control Method for a Current-Source Active power Filter", IEEE Transactions on Industrial Electronics, Vol. 50, No. 2, pp. 313-321, 2003.

[6] H. Akagi, A. Nabae and S. Atoh, "Control Strategy of Active Power Filter using Multiple Voltage-Source PWM Converters", IEEE

Transactions on Industry Applications, Vol. 22, No. 3, May-June 1986, pp460-465.

[7] H. Sasaki and T. Machida, "A New Method to Eliminate AC Harmonic Current by Magnetic Flux Compensation", IEEE Trans. Power Appl. Syst., Vol. PAS-90, No. 5, pp. 2009-2019, 1971.

[8] Ali Ajami, Seyed Hossein Hosseini, " Implementation of a Novel Control Strategy for Shunt Active Filter" ECTI Transaction on Electrical Engg., Electronics, and Communications Vol. 4, NO. 1, Feb 2006.

[9] H. Abaali, M. T. Lamchich, and M. Raoufi. " Shunt Power Active Filter Control under Non Ideal Voltages Conditions" International Journal of Information Technology Vol. 2. Nov. 3 2005 ISSN 1305-239X.

[10] S. Kumar Jaim, P.Agarwal, and H. O. Gupta, "A Control algorithm for compensation of customer-generated harmonics and reactive power", IEEE Trans. On Power Delivery, Vol. 19, no. 1, pp. 357-366, Jan 2004.

[11] Johann F. Petit, Hortensia Amaris and Guillermo Robles, "Control Schemes for Shunt active filters to mitigate harmonics injected by inverter-fed motors". Unpublished.

[12] M. P. Kazmierkowski and M. A. Dzieniakowski, "Review of Current regulation techniques for three-phase pwm inveretes", Industrial Electronics, Control and Instrumentation, 1994. IECON '94. 20th International Conference. Vol. 1, pp. 567-575, Sept 1994.

[13] H. Fujita, H. Akagi, " The Unified Power Flow Conditioner: The Integration of series and shunt Active Filters", IEEE Transaction on Power Electronics, Vol. 13, No. 2, pp. 315-322, 1998.

[14] N. G. Hingorani, L. Gyugyi, "Understanding Facts: Concepts and Technology of Flexible AC Transmission Systems", Wiley-IEEE Press, 1999.

[15] Helder J. Azevedo, Jose M. Ferreira, Antonio P. Martins, Adriano S. Carvalho, "Direct Current Control of an Active Power Filter for Harmonic Elimination, Power Factor Correction and Load Unbalancing Compensation", Unpublished.

◈◈◈

AN APPLICATION OF MATLAB/SIMULINK FOR SPEED CONTROL OF DC SERIES MOTOR USING BUCK CHOPPER

Joy Bandhyapadhyay, Arghya Mukherjee & Suman Ghosh

Electrical Engg. Dept., Dr. B.C Roy Engg. College, Durgapur,India
E-mail : joybndpdh06@gmail.com, arghyamu@gmail.com, eternalsomu@gmail.com

Abstract - In this paper, a strategy, in view of practical simplicity, is implemented to provide a somehow fast and accurate control over non-linear effects of load/switching dynamics. This paper investigates speed control of a dc series motor based on a simulation environment developed within Simulink and Matlab for evaluation of the speed/current control of a prototype series DC motor. It has been shown here the use of buck chopper which paves the way of controlling also torque speed characteristics of dc series motor as well as field current, electromechanical torque .This provides paper provides fast and accurate response alongside with protection to the buck chopper.

Keywords - *DC series motor; buck converter; snubber circuit ; MATLAB/Simulink*

I. INTRODUCTION

With the help of modern software technologies industrial applications have become much more relevant and important in the field of engineering. The Matlab software is a highly versatile and efficient computing language that has been developed by Mathwoks Matlab developers. Not only extremely high, efficient programs can be done here which is much more user friendly, but very complex circuits can also be developed with the help of Simulink. DC motor's speed can be controlled by field resistance, armature resistance and armature voltage methods. But in this paper a technique of drive has been used for dc motor's speed control that is buck chopper. Buck chopper can be implemented by combination of IGBT and Freewheeling diode as well as GTO thyristor and freewheeling diode. The circuits have all been developed here with the help of tools available in simulink [1] and simpowersystems [2] software packages. In this paper dc series motor has been considered as sample. In case of dc series motor we know that its characterised by high torque-low speed operation. So the aim of this paper is to make sure that speed, torque, armature current ,voltage become constant even if load goes on increasing.

II. BUCK CHOPPER

It consists a set of diode and switch along with a low-pass LC filter. Using a proper duty cycle control over the switches, it is possible to control the output voltage. Here, a diode and a switch are needed to implement a complete buck converter operation. From basic power electronics, buck converter may be in either continuous conduction mode (CCM) or discontinuous conduction mode(DCM).The operation of buck converter be it in CCM or in DCM depends upon the load resistance *Rl*, motor's inductance *L*, pulse duration *D*, and switching frequency *1/Ts*. Flywheel is used to reduce the peak demand by the motor in case of heavy loads .A buck converter has its design similar to the step-up boost converter ,and like the boost converter it is a switched-mode power supply that uses two switches (a transistor and a diode), an inductor and a capacitor. It is a step-down DC to DC converter. The linear regulators such as a 7805 can be used to reduce the voltage of a DC supply but linear regulators waste energy as they operate by dissipating excess power as heat. Buck converters, in the other way, can be effectively efficient (95% or higher for integrated circuits), making them useful for tasks such as converting the 12–24 V typical battery voltage in a laptop down to the few volts needed by the processor. The operation of the buck converter is fairly simple, with an inductor and two switches (usually a transistor and a diode) that control the inductor. It alternates between connecting the inductor to source voltage to store energy in the inductor and discharging the inductor into the load

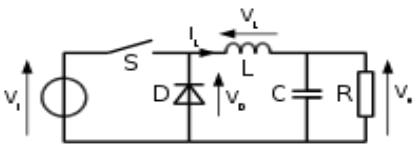


Fig 1. BUCK CHOPPER

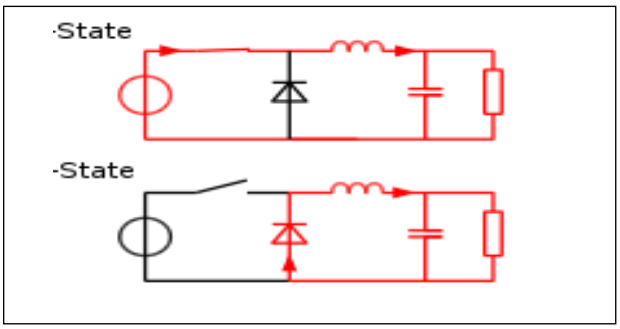


Fig 2. BUCK CHOPPER IN ON-OFF POSITION

III. DC SERIES MOTOR

The DC series motor provides high starting torque and is able to move very large shaft loads when it is first energized.

Since the series field winding is connected in series with the armature, it will carry the same amount of current that passes through the armature. For this reason the field is made from heavy-gauge wire that is large enough to carry the load. Every DC motor is primarily described by the following equations. The Counter EMF is proportional to the machine speed.

$$E = K_V\omega \qquad (1)$$

K_V is the voltage constant and ω is the machine speed. In a separately excited DC machine model, the voltage constant K_V is proportional to the field current I_f:

$$K_V = L_{af}I_f, \qquad (2)$$

where L_{af} is the field-armature mutual inductance. The electromechanical torque developed by the DC machine is proportional to the armature current I_a.

$$T_{emt} = K_TI_a, \qquad (3)$$

Where K_T is the torque constant. The sign convention for T_{emt} and T_L is T_{emt}, $T_L > 0$: Motor mode T_{emt}, $T_L < 0$: Generator mode. The torque constant is equal to the voltage constant.

$$K_V = K_T. \qquad (4)$$

The mechanical part computes the speed of the DC machine from the net torque applied to the rotor. The speed is used to implement the CEMF voltage E of the armature circuit. The mechanical part is represented by Simulink blocks that implement the equation

$$J\frac{dw}{dt} = T_{emt} - B_mw - T_L - T_f \qquad (5)$$

where J = inertia, B_m = viscous friction coefficient, and T_f = Coulomb friction torque. A DC motor is fed by a DC source through a chopper which consists of IGBT and a free-wheeling diode.

VI. SPEED CONTROL

Speed control is achieved by adjusting the field Ampere-turns. It can be done by the following methods :(a) flux control method−

(i) by field diverters that is a low ohmic value variable resistance and high current carrying capacity connected in parallel to the field winding.

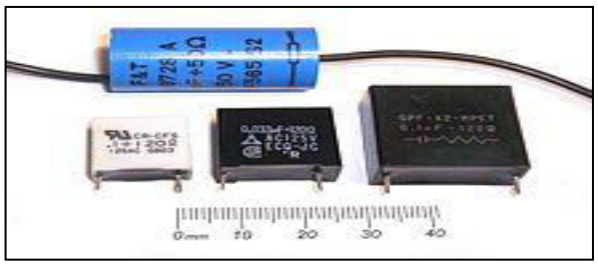
(ii) by armature diverters that is variable resistance connected in parallel to armature winding.

(iii) tapped field control that is flux changing is obtained by using different tappings of the field windings by means of a rotary switch connected to the field tappings.

(iv) Paralleling field coils

(b) variable resistance in series that is normally not used due to high power loss and power input to the armature will become small .

V. RC SNUBBER



A simple snubber uses a small resistor (R) in series with a small capacitor(C). This combination can be used to suppress the rapid rise in voltage across a thyristor, preventing the erroneous turn-on of the thyristor; it does this by limiting the rate of rise in voltage (dV/dt) across the thyristor to a value which will not trigger it. Snubbers are also often used to prevent arcing across the contacts of relays and switches and the electrical interference and welding/sticking of the contacts that can occur. An appropriately-designed RC snubber can be used with either DC or AC loads. This sort of snubber is commonly used with inductive loads such as electric motors. The voltage across a capacitor cannot change instantaneously, so a decreasing transient current will flow through it for a small fraction of a second, allowing the voltage across the switch to increase more slowly when the switch is opened. While the values can be optimised for the application, a 100 ohm non-inductive resistor in series with a 100 nanofarad, or larger, capacitor of appropriate voltage rating is usually

effective. Determination of voltage rating can be difficult owing to the nature of transient waveforms; the actual rating can be determined only by measuring temperature rise of the capacitor. This type of snubber is often manufactured as a single component.

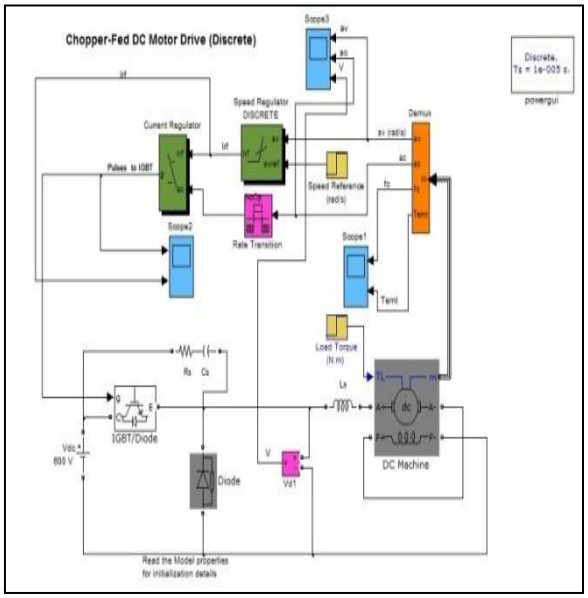


Fig 4. dc series motor fed by buck converter

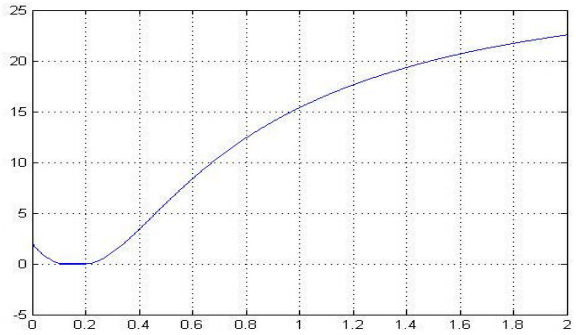


Fig 5. speed vs time

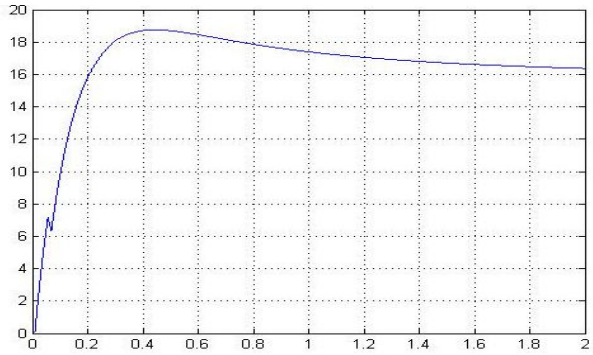


Fig 6.Armature current vs time

VI. CIRCUIT DESCRIPTION

A DC motor is fed by a DC source through a chopper which consists of IGBT(Insulated Gate Bipolar Transistor) and a free-wheeling diode. The motor drives a mechanical load characterized by inertia J, friction coefficient B, and load torque TL. The motor uses the discrete DC machine provided in the Machines library[2].The hysteresis current regulator compares the sensed current with the reference and generates the trigger signal for the transistor to force the motor current to follow the reference. The speed control loop uses a proportional-integral controller which produces the reference for the current loop. The Rate Transition block is used to transfer data to the input of a block from the output of a block, both operating at different rates[2]. Use the block parameters to trade data integrity and deterministic transfer for faster response or lower memory requirements.

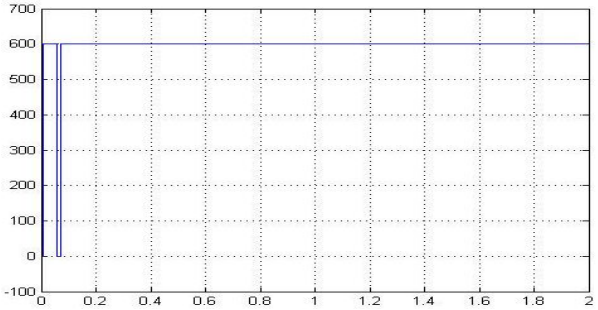


Fig 7.voltage vs time

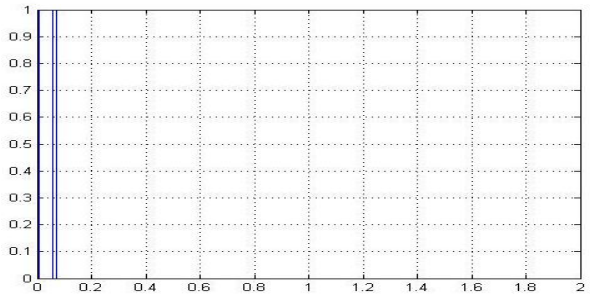


Fig 8. Pulses of IGBT vs time

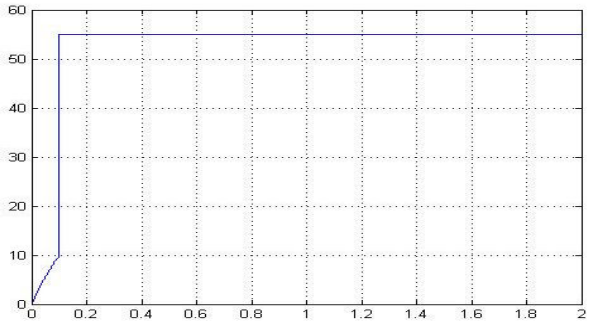


Fig 9.Reference current vs time

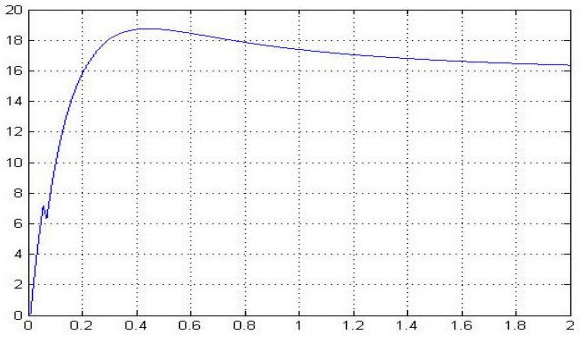


Fig 10.Field current vs time

VII. OBSERVATION TABLES:

A. SPEED Vs TIME

B. ARMATURE CURRENT(Ia) Vs TIME(1)

A> w(rad/s)	t(s)
2.0	0.0
0.0	0.1
0.0	0.2
2.0	0.3
4.0	0.4
12.5	0.8
17.5	1.2
20.5	1.6
22.5	2.0

B> Ia(A)	t(s)
0.0	0.0
7.0	0.098
6.2	0.1
16.0	0.2
18.5	0.4
17.99	0.8
17.0	1.2
16.4	1.6
16.2	2.0

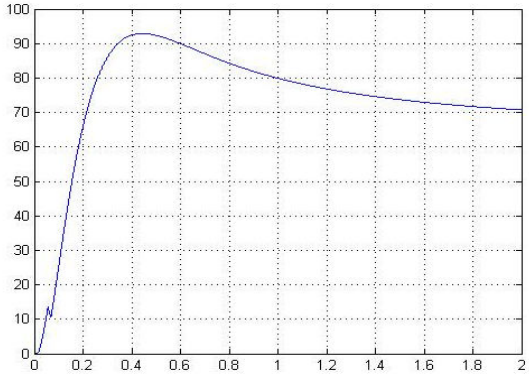


Fig 11. Electromechanical torque vs time

Here mechanical torque input and reference speed are applied in step form from 0.0 to 40.0 Nm and 2.0 to 40.0 radians/second, the motor used has rating: 250HP, 500V, 1750R.P.M, field voltage 300V.As evident from table along with time as load on motor increases the characteristic speed of dc series motor rises after a drop and finally becomes constant(fig 2), so does the armature current(fig 3). Here a discrete PI controller is used for speed regulator. As evident from field current vs time and electromechanical torque vs time graphs both after linearly rising to 18.5A settle down at 16.2A and 92.5Nm to 70.1Nm respectively. So even if load goes on increasing the speed of dc motor gets settled down .

VIII. ABBREVIATIONS AND ACRONYMS

- Discontinuous conduction mode(DCM).
- continuous conduction mode(CCM)
- CEMF-Counter Electromotive force
- E-emf
- Ke-voltage constant
- w-machine speed,
- L_{af} - the field-armature mutual inductance,
- If-field current,
- Ia-armature current,
- Kt-torque constant,
- Temt-electromechanical torque,
- Tl-load torque,
- J = inertia,
- B_m = viscous friction coefficient, and
- T_f = Coulomb friction torque,
- t(s)-time in seconds

IX. CONCLUTION

The MATLAB/SIMULINK software can be not only used to simulate and make programs related to powersystems,communications,controlsystems,drives but can be used to improve their performances which can then be applied in terms of industrial terms circuits related are used to generate electrical power in power plants and provide mechanical work in industries. Today it's used to train up students and even academic professionals in order to understand powersystem behaviour in steady state conditions, ability to simulate transient in power system, transient stability

improvement including power electronics, control systems. The future scope of this paper lies in its hardware implementation as well as its modification scope lies in software part like with the help of PID controller technique, Fuzzy logic, Fuzzy GA logic.

ACKNOWLODGEMENT

It is with our deepest gratitude and appreciation which we offer to our respected Professor Sushanta Dutta of Dr.B.C.Roy Engineering College whos advice on the making of this paper has been of great help.

REFERENCES

[1] SIMULINK, Model-based and system-based design, using Simulink, Math Works Inc., Natick, MA, 2000.

[2] SimPowerSystems for use with Simulink, users guide, MathWorks Inc., Natick, MA, 2002.

[3] F. Antritter, P. Maurer, and J. Reger, "Flatness based control of abuckconverter driven dc motor," Proc. 4th IFAC Symposium onMechatronic Systems, vol. 4, Part 1, 2006.

[4] Boldea and S.A. Nasar, Electric Drives. Boca Raton, FL: CRC PressLLC, 1999.

[5] J. Linares-Flores and H. Sira-Ramirez, "A smooth starter for a dcmachine: a flatness based approach," Proceedings of 1st InternationalConference on Electrical and Electronics Engineering and X Conference on Electrical Engineering, 2004, pp. 589-594.

[6] J. Linares-Flores and H. Sira-Ramirez., "Dc motor velocity control through a dc-to-dc power converter," Proceedings of 43rd Conference.

[7] B. Eskandari, H. Valizadeh Haghi, M. Tavakoli Bina and M.A. Golkar "An Experimental Prototype of Buck Converter FedSeries DC Motor Implementing Speed and Current Controls" International Conference on Computer Applications and Industrial Electronics (ICCAIE 2010),December5-,2010,KualaLumpur,Malaysia.

[8] R. W. Erickson, Fundamentals of Power Electronics, NewYork: Chapman and Hall May 1997.

[9] L. Guo, "Implementation of digital PID controllers for dc-dc converters using digital signal processors," Proceedings of IEEE EIT, 2007, pp.306-311.

◈◈◈

CLUSTER : A Matlab GUI Package for Data Clustering

Sudip Poddar, Anirban Mukhopadhayay
Department of Computer Science and Engineering, University of Kalyani, Kalyani – 741235, WB, India
E-mail: sudippoddar2006@gmail.com, anirban@klyuniv.ac.in

Abstract - The result of one clustering algorithm can be very different from that of another for the same input dataset as the other input parameters of an algorithm can substantially affect the behaviour and execution of the algorithm. Validity measures can be used to find the partitioning that best fits the underlying data (to find how good the clustering is). In most realistic applications, this analysis can be visualised using simple Computer-Aided-Design (CAD) package *(*such as GUIDE tools within MATLAB) specifying various constraints. This paper describes an application (CLUSTER) developed in the Matlab/GUI environment that represents an interface between the user and the results of various clustering algorithms. The user selects algorithm, internal validity index, external validity index, number of clusters, number of iterations etc. from the active windows. In this Package we compare the results of k-means, fuzzy c-means, hierarchical clustering and multiobjective clustering with support Vector machine (MocSvm). This paper presents a MATLAB Graphical User Interface (GUI) that allows the user to easily "find" the goodness of a cluster and immediately see the difference of those algorithms graphically. Matlab (R2008a) Graphical User Interface is used to implement this application package.

Keywords - *Clustering;Validity Index; Matlab; Graphical user Interface; CAD; Interface;*

I. INTRODUCTION

A graphical user interface provides the user with a familiar environment for an application. This environment contains pushbuttons, toggle buttons, lists, menus, text boxes, and so forth, all of which are already familiar to the user, so that he or she can concentrate on using the application rather than on the mechanics involved in doing things. However, GUIs are harder for the programmer because a GUI-based program must be prepared for mouse clicks (or possibly keyboard input) for any GUI element at any time. Such inputs are known as events, and a program that responds to events is said to be event driven. The three principal elements required to create a MATLAB Graphical User Interface are:-

1. Components: - Each item on a MATLAB GUI (pushbuttons, labels, edit boxes, etc.) is a graphical component. The types of components include graphical controls (pushbuttons, edit boxes, lists, sliders, etc.), static elements (frames and text strings), menus, and axes.
2. Figures: - The components of a GUI must be arranged within a figure, which is a window on the computer screen.
3. Callbacks: - Finally, there must be some way to perform an action if a user clicks on a button with mouse or types information on a keyboard. A mouse click or a key press is an event, and the MATLAB program must respond to each event if the program is to perform its function. For example, if a user clicks on a button, that event must cause the MATLAB code that implements the function of the button to be executed. The code executed in response to an event is known as a call-back. There must be a callback to implement the function of each graphical component on the GUI.

Clustering is a popular unsupervised pattern classification technique which partitions the input space of *n* objects into *K* regions based on some similarity/dissimilarity measure.

- The value of K may or may not be known a priori.
- Output of a clustering technique is a K × n matrix $U = [U_{ki}]$. U_{ki} denotes the membership degree of i^{th} object to the k^{th} cluster.
- For crisp clustering, $U_{ki} \in \{0,1\}$ and for fuzzy clustering, $0 < U_{ki} < 1$.

In this article we have designed a MATLAB/GUI package called CLUSTER that implements different clustering algorithms and also computes the values of different cluster validity indices. The results are presented to the user in graphical and tabular forms. Here we implemented k-means, fuzzy c-means, hierarchical clustering and multiObjective clustering with support Vector machine (MocSvm) clustering algorithms in this Application package.

The rest of the article is organized as follows. The next section gives short descriptions of different clustering algorithms included in the package. Cluster validity indices are described in Section III. We have demonstrated the use of the application package (CLUSTER) in Section IV. Finally, Section V concludes the article.

II. CLUSTERING ALGORITHMS

In this paper following clustering algorithms have been implemented.

A. K-Means

K-means [1][2], is one of the simplest unsupervised learning algorithm that solves the well known clustering problem. This algorithm aims at minimizing an ***objective function***, in this case a squared error function.

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \| X_i^{(j)} - C_j \|^2 \qquad (1)$$

Where $\| X_i^{(j)} - C_j \|$ a chosen distance measure between a data point $X_i^{(j)}$ and the cluster center C_j ,is an indicator of the distance of the n data points from their respective cluster centers. K-means minimizes the global cluster variance J to maximize the compactness of the clusters. It has been shown that the k-means algorithm may converge to values that are not optimal.

B. Fuzzy C-Means

Fuzzy C-means (FCM) [3][4], is a method of clustering which allows one data point to belong to two or more clusters with different membership degrees. This method is frequently used in pattern recognition. It is based on minimization of the following objective function

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m \| X_i - C_j \|^2 \quad 1 \le m < \alpha \qquad (2)$$

where $\boldsymbol{m}$ is any real number greater than 1, $\boldsymbol{u_{ij}}$ is the degree of membership of $\boldsymbol{x_i}$ in the cluster $\boldsymbol{j}$, $\boldsymbol{x_i}$ is the $\boldsymbol{i}^{th}$ of d-dimensional measured data, $\boldsymbol{C_j}$ is the d-dimension center of the cluster, and $\|*\|$ is any norm expressing the distance between any measured data and the center.

FCM clustering usually performs better than K-means clustering for overlapping clusters and noisy data. However, this algorithm may also stuck at local optima. Both the K-means and FCM algorithms are known to be sensitive to outliers.

C. Hierarchical Clustering

In Hierarchical clustering [5][6], the clusters are generated in a hierarchy, where every level of the hierarchy provides a particular clustering of the data, ranging from a single cluster (where all the points are put in the same cluster) to n clusters (where each point comprises a cluster). Hierarchical clustering may be either agglomerative or divisive.

Agglomerative clustering techniques begin with singleton clusters, and combine two least distant clusters in each iteration. Thus in each iteration two clusters are merged, and hence the number of clusters reduces by one. Divisive clustering just follows the reverse process, i.e., it starts from a single cluster containing all the points. At each step, the biggest cluster is divided into two clusters until the target number of clusters is achieved.

D. MultiObjective Clustering With Support Vector Machine

In the combined approach, named as MOCSVM [7], each non-dominated solution is given equal importance and a fuzzy majority voting technique is applied. This is motivated by the fact that due to the presence of training points, supervised classification usually performs better than the unsupervised classification or clustering. Here we have exploited this advantage while selecting some training points using fuzzy voting on the non-dominated solutions produced by the multiobjective fuzzy clustering. The fuzzy voting technique gives a set of points which are assigned to some clusters with high membership degree by most of the non-dominated solutions. Hence these points can be thought to be clustered properly and thus can be used as the training points of the classifier. The remaining low-confidence points are thereafter classified using the trained classifier.

Here, the steps of Mocsvm procedure are discussed in detail.

III. PROCEDURE OF MOCSVM

1. Apply multiobjective clustering on the given data set to obtain a set $N = \{S_1, S_2, \ldots, S_N\}$, $N \le P$, (P is the population size) of non-dominated solution strings consisting of cluster centers.
2. Using Eq. (3), compute the fuzzy membership matrix $U^{(i)}$ for each of the non-dominated solutions S_i, $1 \le i \le N$.
3. Reorganize the membership matrices to make them consistent with each other i.e., cluster j in the first solution should correspond to cluster j in all the other solutions.
4. Mark the points whose maximum membership degree (to cluster j, $j \in \{1, 2... K\}$) is greater than a membership threshold α ($0 \le \alpha \le 1$), for at least βN solutions, as training points. Here β ($0 \le \beta \le 1$) is

the threshold of the fuzzy majority voting. These points are labelled with class *j*.

5. Train the multi-class SVM classifier (i.e., *K* one-against-all two-class SVM classifiers, *K* being the number of clusters) using the selected training points.
6. Predict the class labels for the remaining points (test points) using the trained SVM classifier.
7. Combine the label vectors corresponding to training and testing points to obtain the final clustering for the complete data set.

III. CLUSTER VALIDITY INDICES

The result of one clustering algorithm can be very different from another for the same input data set as the other input parameters of an algorithm can substantially affect the behaviour and execution of the algorithm. The main objective of a cluster validity index is to validate a clustering solution, i.e., to find how good is the clustering. Validity measures can be used to find the partitioning that best fits the underlying data. Beyond 3-dimensional space, it is not possible to visualize the clustering result in the feature space. Therefore cluster validity measures can effectively be used to compare the performance of several clustering techniques, specially for high dimensional data. There are mainly two types of cluster validity indices: *external* and *internal*.

A. *External Cluster Validity Indices*

External validity measures are used to compare the resultant clustering solution with the true clustering of data. These indices are very useful for comparing the performance of different clustering techniques when the true clustering is known. In this application package (CLUSTER), the external cluster validity indices used are Minkowski index [14], Adjusted Rand index [15] and percentage of correctly classified pairs [16]. For adjusted rand index and percentage of correctly classified pairs, higher value indicates better matching of true clustering and clustering result provided by an algorithm. On the other hand, lower value of Minkowski index indicates better matching between the true and obtained clustering.

B. *Internal Cluster Validity Indices*

Internal validity indices evaluate the quality of a clustering solution in terms of the geometrical properties of the clusters, such as compactness, separation and connectedness. In this article we have implemented some widely used internal cluster validity indices, which are J [17], Davies-Bouldin (DB) index [18], Dunn index [19], Xie-Beni (XB) index [20], I index [21] and Silhouette index [22]. Validity measures that are function of cluster compactness and separation, such as DB, Dunn, XB, I etc. can be used to determine the number of clusters also. In order to achieve proper clustering the values of Dunn Index, Silhouette Index and I Index are to be maximized. On the contrary the values of J Index, Davies-Bouldin Index, and Xie-Beni Index are to be minimized in order to achieve proper clustering.

IV. DEMONSTRATION OF CLUSTER PACKAGE

Matlab/GUI [8] is a programming tool that is used to develop a front-end for a software application. Communication of the user with the programming application used to take place through commands issued by use of keyboard. Today that communication usually takes place by use of a mouse and a graphical interface i.e. via interactive cursors, drop-down menus, slide-bars etc. In this Matlab/GUI application package the used Graphical Components are pushbutton, Radio Button, Edit Box, Static Text Box, Pop-Up menu, Toggle Button, Table, Axes, Panel, Button Group, Labels, and Menus etc. Thousands of Matlab/GUI codes [9][10][11][12] are written in the background of each window.

In this section we describe the proposed Matlab/GUI based application package CLUSTER. For the shortage of allowed pages here we explain the application of Matlab/GUI on a single Dataset of Heart Disease [13] only. The Dataset should be real and no missing value is allowed. Actual number of clusters of the selected Dataset is 2.

A. *Initial Window*

After starting the application the initial window (Homepage) is appeared as shown in Figure 1.

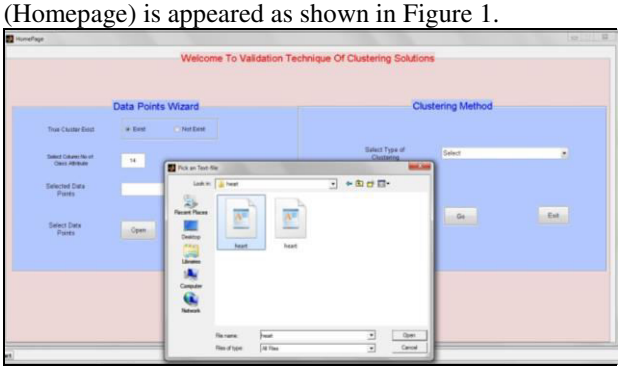


Figure 1: Homepage of CLUSTER package

After selecting the Dataset; If the dataset contains true clustering information then select the exist field in the Toggle Button and also specify the class attribute column number. Thus the class attribute is separated from the Dataset and saved in a different 'text' file into a new folder. Next select the algorithm and Open the

corresponding window. Here we select KMeans clustering algorithm for demonstration.

A. KMeans Window

The KMeans window looks like Figure 2.

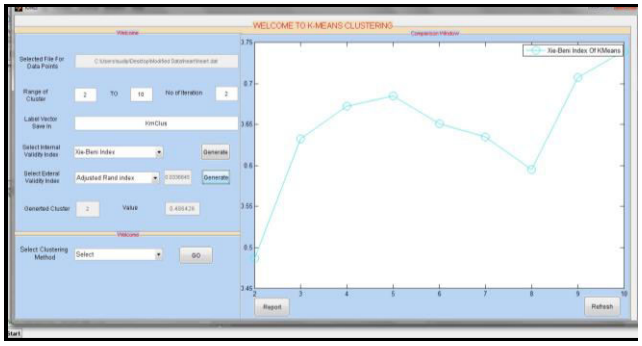


Figure 2: Kmeans clustering window

When this window is opened the Dataset is automatically saved in the Edit box of the selected Data point field. Next specify the range of cluster, number of iterations, the directory where the generated label vector will be saved, select internal validity index. Default value is given to some field i.e. number of iteration is 2, Label vector saved directory (Here KmClus) etc. There is also an option for selecting the external validity index. If true cluster does not exist then all the options for external validity index becomes invisible. After clicking the Generate button of internal validity index, number of generated cluster and corresponding value will be shown in the appropriate boxes and the graph will be drawn. For each cluster the corresponding index value is marked with a marker on the graph. Selecting internal validity index and algorithm is shown in the graph with the help of legend in different color. Scaling of X-axis is defined according to the value used in Range of cluster. Y-axis is scaled according to the generated value.

Following logic is used behind the generated value (maximum or minimum depending upon Internal validity index used) and corresponding number of generated cluster is explained below:-

- Suppose we take Range of cluster from 2 to 10 and number of iterations 5.
- We used Dunn Index, Silhouette Index, I Index, J Index, Davies-Bouldin Index, and Xie-Beni Index as Internal Validity Index. Lager Value gives better results for Dunn Index, Silhouette Index, I Index and smaller value gives better results for J Index, Davies-Bouldin Index, and Xie-Beni Index. Suppose we select Xie-Beni Index.
- For Iteration Number 1
  - For each number of clusters an Index value is generated. Here 9 Index (for Xie-Beni) value is generated for the given range of number of clusters. In this case Minimum value gives better result. So we take the minimum one among the 9 Index values. We also save the corresponding generated label vector.

At each iteration, we get one minimum index value and generated label vector of each index value. Thus for 5 iterations we get 5 minimum values and 5 different label vector. Among these we take the minimum one and the corresponding label vector. The Generated label vector is saved in assigned directory (here KmClus).

Thus we also calculate the number of generated clusters from the generated label vector. Number of generated cluster and corresponding value also saved in the Report table. This table is also seen by clicking the Report button. The Graph is also refreshed by clicking the Refresh button. Value of External validity index will be generated after clicking the Generate button; this value appears in the appropriate box and is saved in Report Table. Here we used Adjusted Rand Index, Minkowski Index, and percentage of correctly classified pairs as External Indices. Next we select the Fuzzy C-Means Clustering Algorithm.

C. Fuzzy C-Means Window

The Matlab/GUI code is written in such a way that at the moment of opening this window all the values which are already set by the user in KMeans window are automatically saved in the corresponding field in this window, except the label vector saved directory (default value given) . Then we click the Generate button of Internal and External validity index. Corresponding values are set in the boxes and also saved in the Report Table. The Graph is drawn automatically with different color. The algorithm and selected internal validity index is shown in the graph below the legend of KMeans. The Fuzzy C-Means Window is shown below in Figure 3.

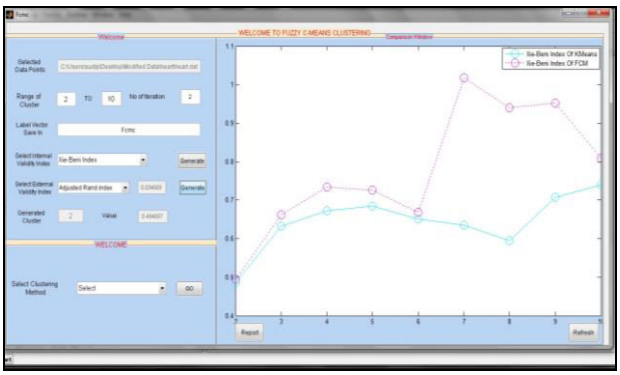


Figure 3:Fuzzy C-means clustering window

Now we go to Hierarchical Clustering window.

D. Hierarchical ClusteringWindow

After opening the Hierarchical clustering window, selected Data points, range of clusters, number of iteration, Internal as well as external validity index field values are set like KMeans window as mentioned before. Here we have two extra fields, one is 'Distance' and another one is 'Method'. Select one of the Distance values among Euclidean, Seuclidean, Cityblock, Mahalanobis, Minkowski, Cosine, Correlation, Spearman, Hamming, Jaccard, Chebychev. Also the various 'Method' are Single, Complete, Average, Weighted, Centroid, Median, Ward. Figure 4 shows the Hierarchical clustering window.

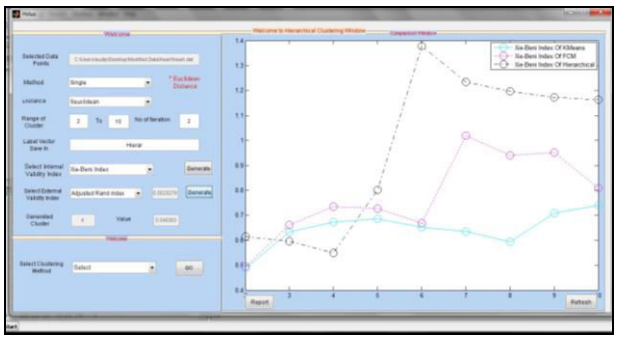


Figure 4: Hierarchical Clustering window

Internal validity index value, number of generated clusters, External validity index value is generated same way as described above. All these values are saved in Report table and the Graph is drawn and marked with a marker in different color. Selected internal validity index and algorithm are shown in the Graph below the previous legend.

Next we open the MocSvm clustering.

E. MocSvm Clustering

The fields which are automatically saved after opening of the window are selected Data Points, Range of Clusters, number of iterations, internal validity index, and external validity index. Default values are given in few fields which are number of generation, population size, Pcrossover, Pmutations, Alpha, Beta, Weight, and the directory to save the Label Vector. Internal validity index value and corresponding label vector are generated by clicking on the generate button of internal validity index. Hence we find the index value and number of generated clusters, external validity index value are also calculated. These values are displayed in appropriate boxes and saved in Report Table. The Graph is drawn, marked with marker and legends are also displayed with different color. If necessary we can refresh the axes with Refresh button. This Graphical window is displayed in Figure 5.

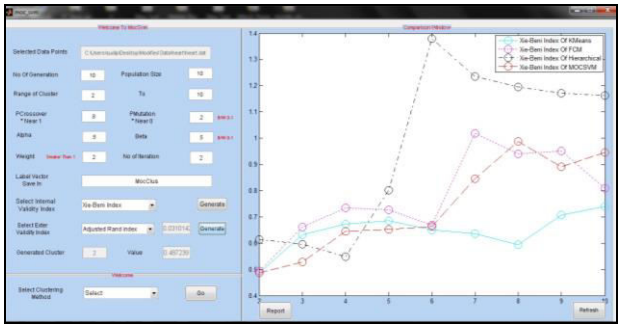


Figure 5: MocSvm clustering window

F. Report Table i.e. Result window

This window is shown in Figure 6.

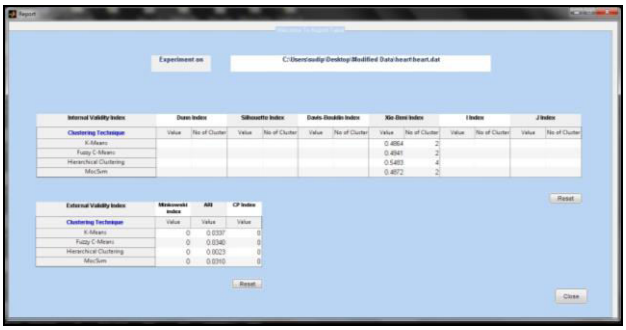


Figure 6: Report Table Window

This window contains two tables, one is internal validity index and another one is external validity index table. Internal validity index table has different algorithms in rows and different validity index in columns. Each Validity index column is divided into two sub-columns, one is Value (maximum or minimum depending in validity Index used) and another one is number of generated clusters. This window is always opened in invisible mode to insert values during the execution of each algorithm window. External validity index table has different algorithms in rows and different validity index in columns. If true clustering does not exist then this table does not appear (invisible). Reset button is used to reset the tables. This window also displays the Dataset on which experiments are done, which is initialized after opening this window. To set the visibility mode off, Close button is used.

G. All Cluster Running Window

For users flexibility the we also develop a window, namely All cluster window. If user wants to run all algorithms with a 'single click', then he/she can do it with this application window. The graphical Matlab/GUI window is shown in Figure 7.

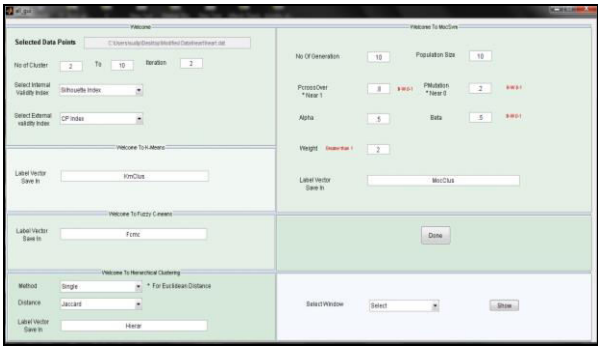


Figure 7: All-Clustering Running window

The selected Data Points field is initialized after opening of this window. All the fields of each algorithm (KMeans, Fuzzy C-Means, Hierarchical Clustering, and MocSvm) are designed in different panels. There is also one common panel of all algorithms which consists of selected Data Points, number of Cluster, Iterations, pop-up menu of internal validity index and external validity index. Default values are given to few fields i.e. iterations, Label Vector Saving Directory, number of Generation, population Size etc. The background code is written in such a way that after clicking on "Done" button it first opens the K-Means Window with invisibility mode. Then sets all the fields of KMeans window (Take the values which are given to the KMeans panel of All Clustering Window). After that it hits on the Generate button of internal validity index and then external validity index. All the tasks of KMeans window are now performed and the values populate the Report Table.

The code is designed in such a way that, Fuzzy C-Means Window, Hierarchical Window and MocSvm windows are opened one after another with invisibility mode. Then the same process takes place as described above but the background Matlab codes are different for calling different window. After running all the windows, Report table is populated with obtained values. Here we used same Dataset of Heart Disease [13] .

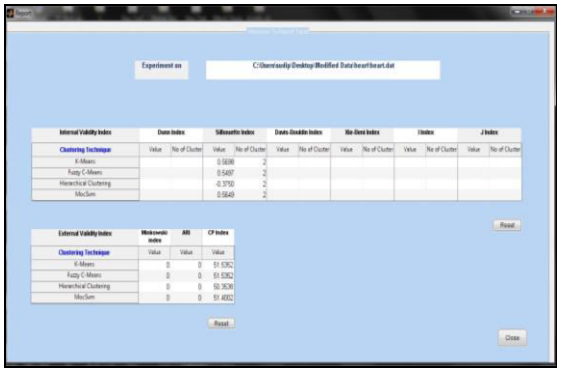


Figure 8: All Clustering Report Table Window

Used internal validity index is Silhouette Index (Maximum Value gives Better results) and external validity index is CP index. The result is shown in Fig. 8.

If user wants to see all other windows as well as Report table then he/she can use the options given in Pop-up menu of "Shown" button.

V. CONCLUSION

In this paper, we have presented an effective and user-friendly MATLAB GUI Application tool for data clustering. We have developed this analytical design tool and software using MATLAB tool boxes in such a way that, if any user wants to insert a new algorithm he/she can do it. He/she can also insert new validity index methods. The Matlab/GUI application package is also available from the authors by request.

REFERENCES

[1] J. B. MacQueen (1967): "Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press, 1:281-297 ., in press.

[2] Andrew Moore: "K-means and Hierarchical Clustering – Tutorial Slides"http:// www2. cs.cmu.edu/ ~awm/tutorials/kmeans.html.

[3] J. C. Dunn (1973): "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", Journal of Cybernetics 3: 32-57. , in press.

[4] J. C. Bezdek (1981): "Pattern Recognition with Fuzzy Objective Function Algoritms", Plenum Press, New York. ,in press.

[5] S. C. Johnson (1967): "Hierarchical Clustering Schemes" Psychometrika, 2:241-254. ,in press

[6] R. D'andrade (1978): "U-Statistic Hierarchical Clustering" Psychometrika, 4:58-67. , in press.

[7] U Maulik, A. Mukhopadhyay and S. Bandyopadhyay, "Combining Pareto-Optimal Clusters using Supervised Learning for Identifying Co-expressed Genes", BMC Bioinformatics, Vol. 10, No. 27, 2009.

[8] The Mathworks, Building GUIs with MATLAB, Natick, MA: The Mathworks, Inc., 1997.

[9] http://www.mathworks.in/matlabcentral/index.html.

[10] http://blinkdagger.com/matlab/

[11] http://www.intelligent-systems.info/classes/ee509/gui.htm

[12] http:// www.youtube.com/watch?v= D_hmws6dwgg

[13] http://archive.ics.uci.edu/ml/datasets/Statlog+%28Heart%29

[14] A. Ben-Hur and G. Isabelle, "Detecting stable clusters using principal component analysis," pp. 159–182, 2003.

[15] K. Y. Yeung and W. L. Ruzzo, "An empirical study on principal component analysis for clustering gene expression data," Bioinformatics, vol. 17, no. 9,pp. 763–774, 2001.

[16] S. Bandyopadhyay, U. Maulik, and A. Mukhopadhyay, Multiobjective genetic clustering for pixel classification in remote sensing imagery," IEEE Trans.Geosci. Remote Sens., vol. 45, no. 5, pp.1506–1511,2007

[17] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms. New York: Plenum, 1981.

[18] D. L. Davies and D.W. Bouldin, "A cluster separationmeasure," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 1, pp. 224–227, 1979.

[19] J. C. Dunn, "Well separated clusters and optimal fuzzy partitions," J. Cyberns., vol. 4, pp. 95–104, 1974.

[20] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, pp. 841–847, 1991.

[21] U. Maulik and S. Bandyopadhyay, "Performance evaluation of some clustering algorithms and validity indices," IEEE Trans. Pattern Anal. Mach. Intell, vol. 24, no. 12, pp. 1650–1654, 2002.

[22] P. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," J. Comp. App. Math, vol. 20, pp. 53–65, 1987.

◈◈◈

A Novel Multi-Planar Band Structure Algorithm : Planar Region Matrix Approach

Deepak Bharti & Abhijit R. Asati
Department of Electrical & Electronics Engineering, Birla Institute of Technology & Science, Pilani, India
E-mail : vipashyana3888@gmail.com, abhijitmicro@gmail.com

Abstract **-** The device simulators are used to extract device characteristic parameters and also for drawing the band diagrams. The band diagram generators available in present simulators are generally unidirectional. In this article, we will present an algorithm which can be used to draw the device band diagram along a user specified plane in either vertical or horizontal direction for a user specified biasing conditions.
This work highlights basics of device level tool development and is useful to the students and device engineers, offering them an interactive and user friendly way to draw band diagram. The algorithm implementation is matrix based and requires GUI development and function development which is performed using MATLAB®.

Keywords- *device simulator, band-diagram, algorithm, MATLAB®, region matrix.*

I. INTRODUCTION

Field of Integrated circuit technology is one of the fastest growing industries in the world. In coming decades, semiconductor fraternity is set to witness life shaping inventions. Many new concepts and devices were proposed during past decades and still, much more are to be explored. But this exploration becomes markedly convenient with the help of a simulator, as physical fabrication and testing of each new device without knowing its physical compatibility is not feasible. One of the important results, which a simulator produces, is a band structure. There are many situations during device analysis when a designer has to analyze the device structure at different planes for numerous parameters such as band structure variables like forbidden gap (E_g), Fermi level (E_f), top valence level (E_v), bottom conduction band level (E_c), potential barrier (V_o), currents, mobility etc.

In this article, we will present a MATLAB® based implementation of an algorithm to draw the device band diagram along a user specified plane in either vertical or horizontal direction for a user specified biasing conditions. This work can be useful to the students as well as to device designers. This tool will be of great help to the students who do not have higher level mathematics background or knowledge of quantum physics. To address the needs of such students, we have developed the algorithm in an interactive and user friendly way. Primary aim of the work is to enable a naïve student to understand and appreciate the relationship between the band structure and external stimulations like doping and biasing. After inspecting the band structure of externally stimulated device a student must learn that how the band structure responds to changes in applied bias or doping levels [1]. Apart from students, our algorithm is also important for device designers, who, many times need band diagram of a device along different planes. Since professional bodies do not share algorithm details, there is no lucid documentation available on specific algorithms to be used for such purposes. Our work gives an important insight into basics of tool development for device level designing. This work can be a guiding path for novices and amateurs in the field of tool developing and help them to proceed in a certain direction to bring forth more innovative ideas and solutions. Also, by incorporating such an algorithm into the already existing tools, versatility and robustness of the tools can be enhanced.

Work presented in this article deals with 2 dimensional (2-D) structures of microelectronic devices. For algorithm description we have used flow charts which are illustrated in a simplified way for ease of the reader.

The section II describes about the terminology and assumptions used in the implementation of the algorithm and Section III describes the details of the algorithm. Section IV discusses the results while Section V contains conclusion.

II. TERMINOLOGY & ASSUMPTIONS

The algorithm development has been done in MATLAB®. For convenience of users we have

developed GUIs (Graphical User Interfaces), which will ensure smooth interaction with the user. Data structures used in the algorithm are mainly arrays (or matrices), simple and easy to access. For enhanced lucidity, algorithm follows modular structure (divided into several functions instead of keeping all functionalities together) which reduces the design complexity. There is a main script file, from which execution control invokes various functions. All functions are nested themselves i.e. execution control in these functions invokes other sub-functions.

In this article, a novel algorithm is presented for multi-planar band diagram plotting using proposed planar region matrix (PRM) approach as described later in section III. The PRM is a matrix of regions coming along a plane. In this approach, we first extract a matrix of regions of input structure coming along a plane, which is specified by the user. Once matrix is extracted, then depending on various combinations of adjacent regions, various mathematical computations are done; using which band diagram is plotted. For starting simplicity we have assumed that all regions are rectangular in shape. If overlap occurs between two regions, then region coming later in the sequence will be the region in the overlapped area. Doping profile is uniform and ionization is 100 %. Junctions are assumed to be of step type. Expressions of junction potential and depletion widths for homo-junctions and hetero-junctions are assumed as in [2]. Some other assumptions are also taken like depletion approximation and applied voltage appears across depletion region [2], [3].

Table I describes various functions used in the algorithm and table II gives list of variables. Table III and table IV deals with materials and types (n-type, p-type, oxide, metal) respectively and their convention considered in the algorithm.

Inputs are provided using an input text file, which will open automatically after welcome GUI. A user can mention the device structure in terms of co-ordinates, material, type, and external stimulation (doping and biasing).

Figure 1 shows the execution control traversing (or functional hierarchy) diagram for the functions, mentioned in table I.

III. ALGORITHM: THE PLANAR REGION MATRIX APPROACH

In earlier sections we have provided brief overview of terminology, functional behavior and execution control. The material constants [4], [5] considered in our algorithm are described in a separate text file (material constant text file) which can be accessed by any function.

The user specifies the data set for each region in input text file, which contains information such as co-ordinates, material, type, and external stimulation (doping and biasing) pertaining to that particular region.

Required calculations then can be performed for given data sets within any function by reading material constant text file. The details of program and several functions used in the algorithm are described sequentially below.

Table – 1 : Functions used in Algorithm

Function	Meaning / comments
main_script_bd.m	Main script file
welcome_gui.m	Welcome GUI
structure_plotter.m	Plots structure of device specified in Input text file
gui_pc.m	This GUI asks user to enter plane information along which BD is desired.
axsreduction.m	Reduces initial axis
regnmatgen.m	Region matrix generator
regnmatred.m	Reduces region matrix
descpngen.m	Description generator
calculator.m	Calculates calc
bdp.m	Band diagram plotter
drp2s.m	Depletion region plotter for semiconductors
drp2m.m	Depletion region plotter for metals
nrp.m	Neutral region plotter

Table -2 : variables used in algorithm

Variables		Meaning / comments
xz,cdnt		xz is specified direction; cdnt is specified co-ordinate.
regnmatrix_axs		regnmatrix_axs is a 1X2 cell;
1.	regnmatrix	Extracted region matrix
2.	act_axs	Actual axes
lmbx, lmbz		Matrix of leftmost bottom x and z co-ordinates
lx, hz		Matrix of length and height of regions
regnmat_m		Modified region matrix
Descpn		Descpn is 1X5 cell; (shown below)
1.	mat2	Matrix of materials (refer to table III)
2.	type2	Matrix of materials (refer to table IV)
3.	conce	Matrix of electron concentration

4.	conch	Matrix of hole concentration
5.	bias	Matrix of bias applied to metals if any
Calc		Calc is a 1X5 cell; (shown below)
1.	Vo	Matrix of contact potentials (semiconductors)
2.	phib1	Matrix of barrier height on left side (metals)
3.	phib2	Matrix of barrier height on right side (metals)
4.	depw	Depletion region width matrix
5.	bias0	Matrix of bias on metals and on semiconductors

TABLE III. MATERIALS USED AND CONVENTION

Material	Convention in Algorithm
Si	1
GaAs	2
InP	3
SiO2	4
Al	5

TABLE IV. TYPES USED AND CONVENTION

Type	Convention in Algorithm
n- type	1
p- type	2
oxide	3
Metal	4

A. Main Script File (main_scrpt_bd)

As shown in the figure 2, the m-file 'main_script_bd' is the main script file (MATLAB program) in which required functions are called in proper sequence as described in figure1.

B. First GUI (welcome_gui)

Welcome GUI as shown in figure 3 serves as the first interface between user and machine. After selecting 2-D and pressing OK a text file will appear for entering device specifications as shown in figure 4.

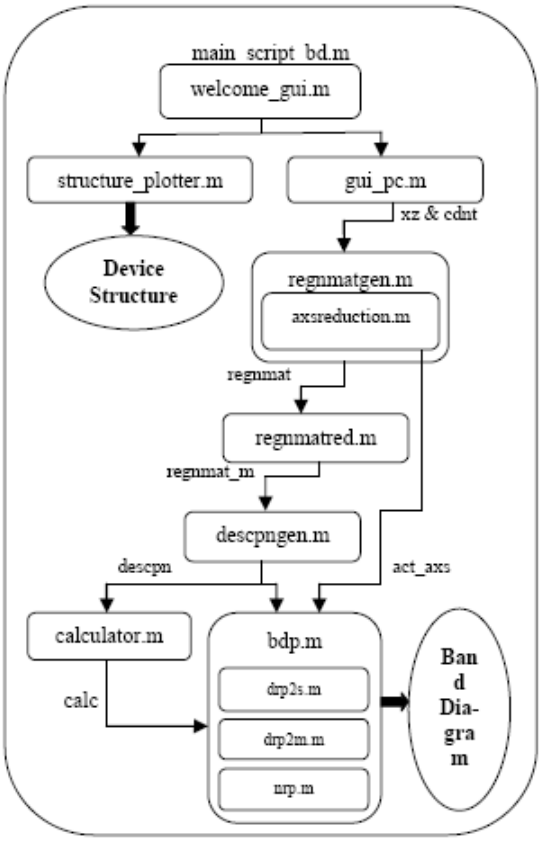


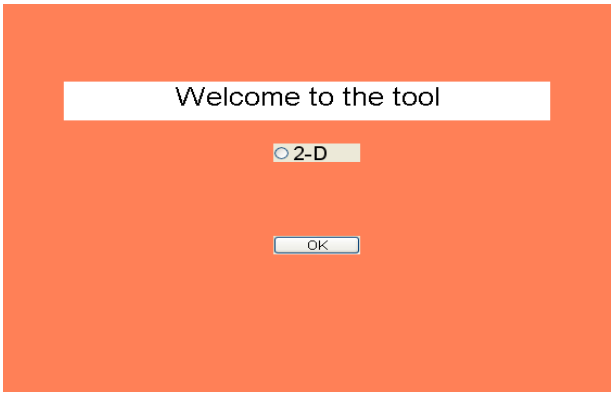
Figure1. Execution Control Path of the Algorithm

```
clc;
all clear;
 % welcome GUI %
welcome_gui;
uiwait;
% function structure_plotter %
structure_plotter;
uiresume
% welcome GUI %
gui_pc;
% Extracting value of plane and co-ordinate
%
fid2=fopen('C:\Documents and
Settings\ert\MyDocuments\dump.txt','r');
B = textscan(fid2,'%f',2,'delimiter', '');
xz=B{1}(1);
cdnt=B{1}(2);
fclose(fid2);

% function regnmatgen %
 regnmatrix_axs=regnmatgen(xz,cdnt);

regnmatrix=regnmatrix_axs{1,1};
act_axs=regnmatrix_axs{1,2};
% function regnmatred %
regnmat_m=regnmatred(regnmatrix);
% function descpngen %
Descpn=descpngen(regnmat_m);
% function calculator %
Calc=calculator(Descpn);
% function bdp %
bdp(Calc,Descpn,act_axs);
```

Fig. 2 : Main Script File (main_script_bd.m)



First GUI (welcome_gui)

In the input text file leftmost bottom x, leftmost bottom z, length and height have their usual meanings. Material and type conventions are provided in the table III and table IV. If material is a semiconductor then user is required to enter doping concentration in multiple of 10^{16} /cm^3. If material is a metal then user can enter applied bias under same column since bias can be applied through metal only.

All entries in a row should be separated by an empty space. After entering device information in desired format save the text file and close it.

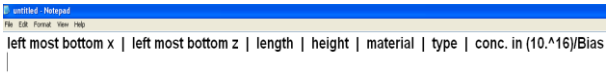


Figure 4: An Empty Input Text File

C. *Device Structure (structure_plotter)*

After closing the saved input file, a 2-D image of device will appear which will be the outcome of the function 'structure_plotter'. For example, we entered the some arbitrary device information in the input text file, as shown in figure 5. After closing the input file a 2D structure will appear as shown in figure 6.

Function 'structure_plotter' opens the input file and reads the first 5 values (lmbx, lmbz, lx, hz, mat2) for each data set until it finds an empty data set. Then it plots all the read values in the form of rectangles.

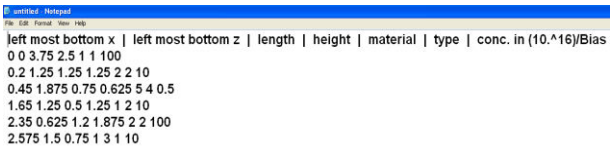


Figure 5. Data Entered In Input Text File

D. *Second GUI (gui_pc)*

In addition to above 2D structure, another GUI also appears simultaneously after closing the input file as shown in figure 7. It appears due to execution of the function 'gui_pc'. This GUI will ask user to enter the plane information in terms of direction and co-ordinate along which band diagram is desired. This GUI offers the user to select plane along x or z.

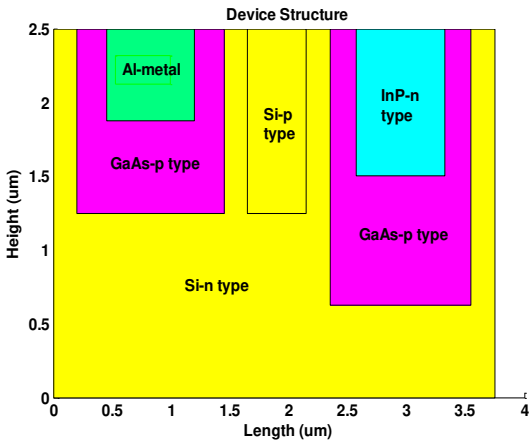


Figure 6. Device Structure

direction, which is an added advantage of the algorithm.

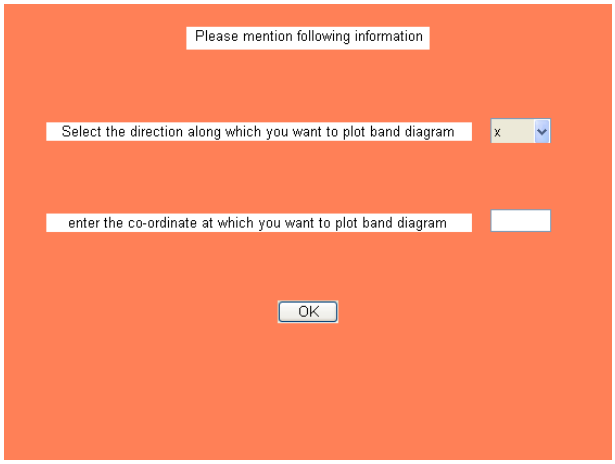


Figure 7. Second GUI (gui_pc)

E. *Region Matrix Generation (regnmatgen)*

Once direction and co-ordinate are selected and OK is pressed then execution control invokes the function 'regnmatgen', which is the most important module of the algorithm, as it extracts the planar region matrix, on which all later calculation will be based.

A brief detail of the logic used to determine planar region matrix is shown in figure 8.

Robustness of the algorithm can be best understood by analyzing the results for various combinations of directions and co-ordinates for various types of devices. All results of such possible combinations for arbitrary device structure considered in the example are shown below in figure 10 in pictorial form for better

understanding of the reader. Arrows in the figure 9 represent various planes possible in the example structure figure 6.

In this example there are total 16 planes possible and hence 16 different region matrices associated with these planes. Numbers written either above or left of an arrow are the content of region matrix along that particular plane.

E.1 Axes Reduction (Sub-function 'axsreduction')

For the user specified plane, the generated region matrix may not contain few regions, hence the axes described in

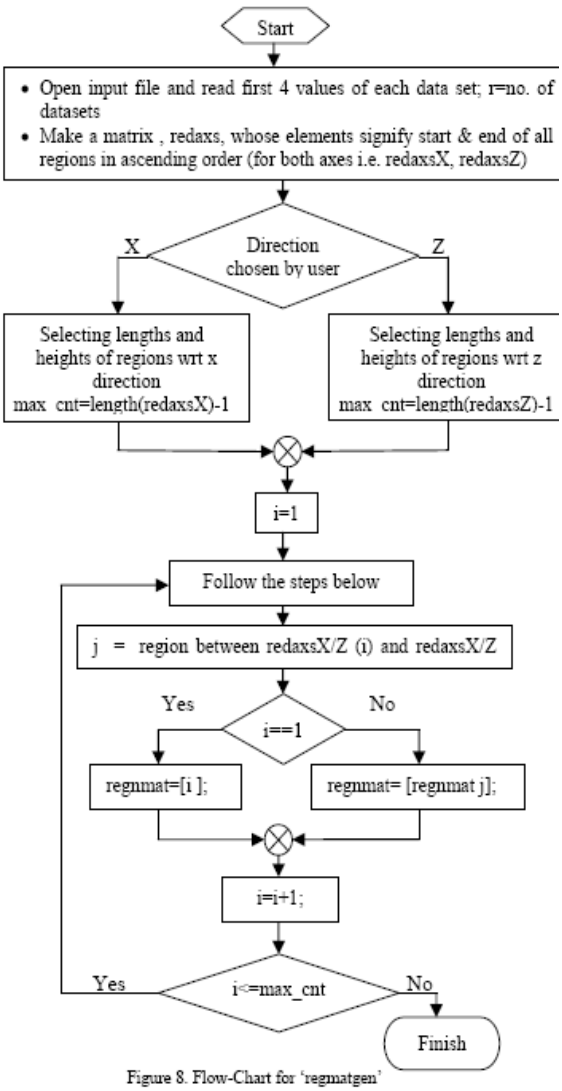


Figure 8. Flow-Chart for 'regmatgen'

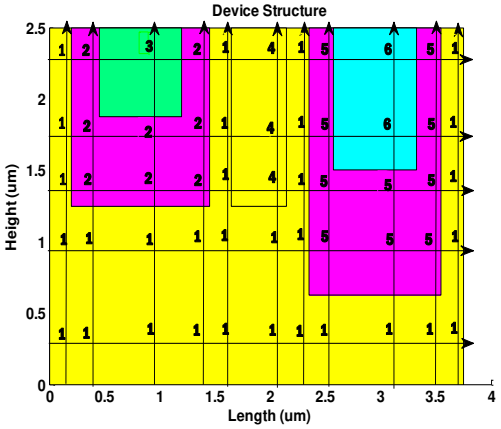


Figure 9. Possible Horizontal & Vertical Region Matrices region matrix can be reduced. Initial axes generated using region matrix generator along x or z axes is indicated in figure 11 by dashed lines. But, actual axes of a plane may not be always same as initial axes. For example, for the first horizontal plane from top, actual axes are same as initial axes (total number of axes points are 12). But actual axes is not same as initial axes for the second horizontal plane from top, which is shown bold in figure 11 (total number of axes points are 10).

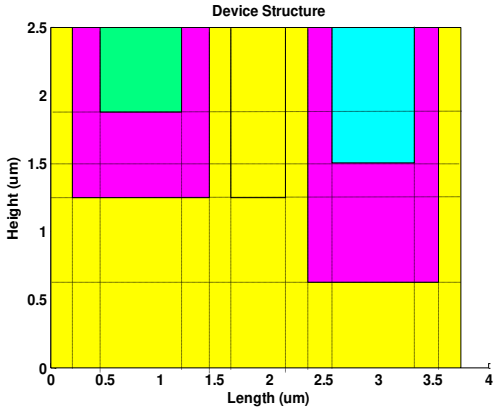


Figure 10. Initial Horizontal & Vertical Axes

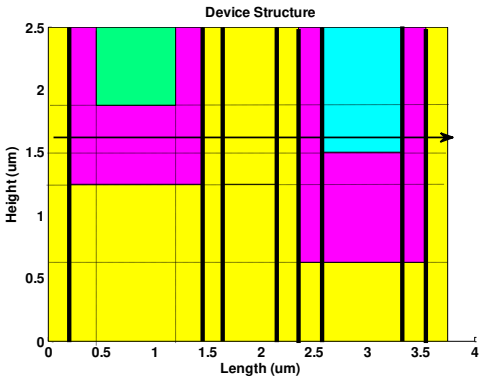


Figure 11. Actual Axes (bold) for Second Horizontal Plane From Top

F. Region Matrix Reduction (regnmatred)

We can identify two or more similar adjacent regions to reduce into a single region using function 'regnmatred'. The resulting matrix is called reduced region matrix. Results of this function can be more clearly observed in figure 12. For better representation, axes are not marked, only horizontal and vertical planes are shown in separate figures. Numbers written either above or left of an arrow represents the reduced region matrix along that plane.

G. Description Generation (descpngen)

Now onwards, all calculations will be done on the reduced region matrix obtained as described in section F. The function 'descpngen' uses reduced region matrix to generate five specification matrices, namely, material (mat2), type (type2), electron concentration (conce), hole concentration (conch) and bias (bias). Procedure used to generate above matrices is shown in the flow chart of figure 13.

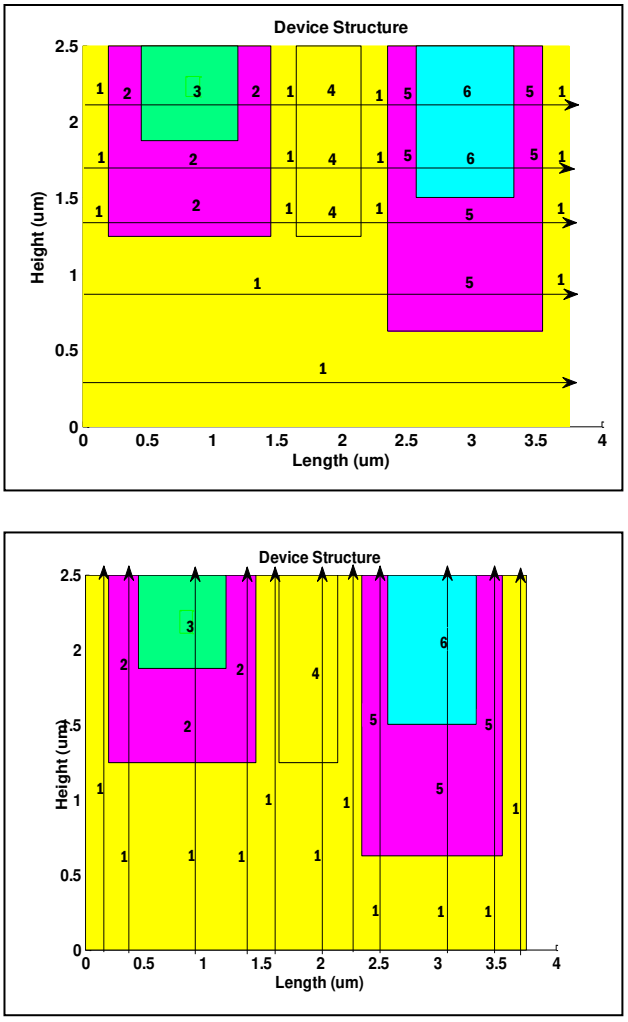


Figure 12. Reduced Region Matrices,(a) Horizontal; (b) Vertical

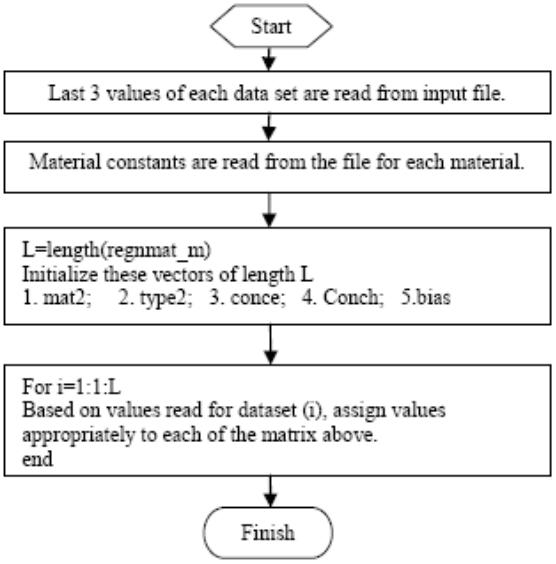


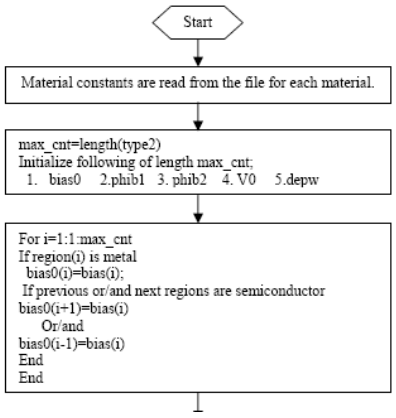
Figure 13. Logic Used For Description Generation

H. Calculations (Calculator)

Results of function 'descpngen', Descpn serve as raw matrices for further mathematical calculations such as height of potential barriers and depletion widths using function 'calculator'. Procedure used is shown in figure 14.

I. Band Diagram Plot (Bdp)

Output of function 'calculator', 'descpngen' and variable 'act_axs' serve as the input to the function 'bdp', that performs the task of plotting the band diagram. In this function differential equations are numerically solved and plotted by MATLAB® compiler. Logic used for 'bdp' is shown in figure 15. This function has three sub-functions, 'drp2s' (depletion region plotter for semiconductors), 'drp2m' (depletion region plotter for metals), 'nrp' (neutral region plotter).



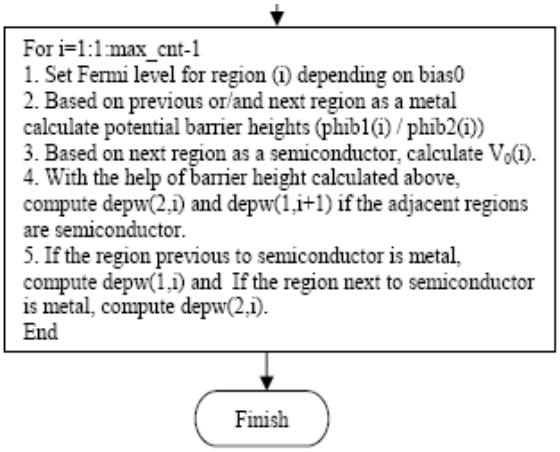


Figure 14. Logic Used For Calculations (calculator)

11. Depletion Region Plotter For Semiconductors (drp2s)

As mentioned earlier, 'drp2s' solves and plots solution of Poisson's equation for semiconductor junctions in depletion region. Its input arguments include adjacent actual axes elements, depletion widths, doping concentration, Fermi level, forbidden energy band, permittivity, difference of Fermi and conduction or valence band level, type and side (it signifies first or second depletion portion of a region for which function is called). The 'drp2s' uses FDM method to solve the differential equations.

I.2 Depletion Region Plotter For Metals (drp2m)

'drp2m' solves and plots solution of Poisson's equation for metal-semiconductor junction in depletion region. Its input arguments are similar to that of 'drp2s' except one inclusion of parameter 'barrier height'. 'drp2m' also uses FDM method to solve the differential equations.

I.3 Neutral Region Plotter (nrp)

'nrp' is called to plot energy levels in neutral portion of semiconductors. Its input arguments are similar to that of 'drp2s'

IV. RESULTS AND DISCUSSION

Once the execution control traverses the complete path shown in figure 1, the desired band diagram is obtained as output. We will show and discuss some of the results for the arbitrary structure figure 6. Planes were selected both in horizontal or vertical directions and band diagrams are plotted using proposed algorithm which produced the output as expected and are shown in figure 16.

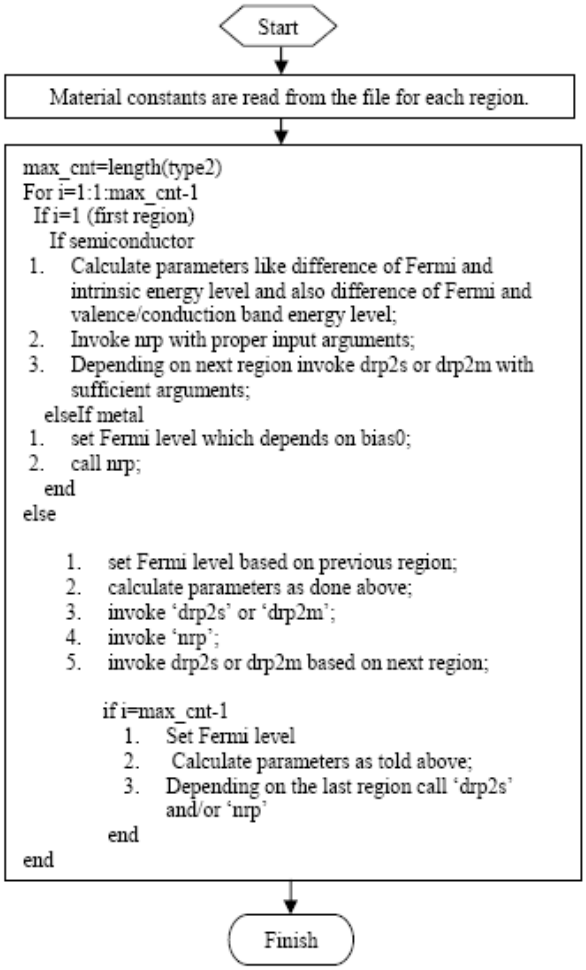
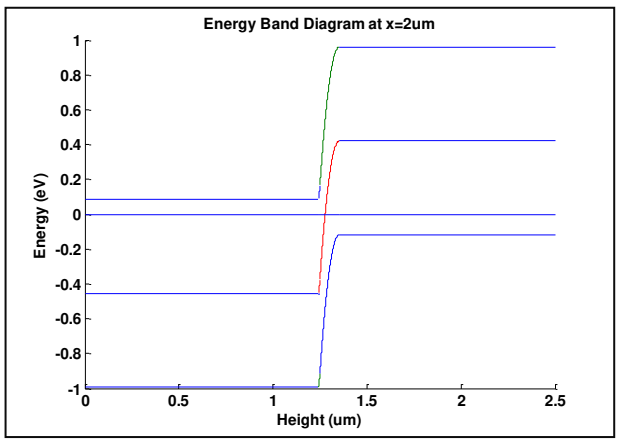
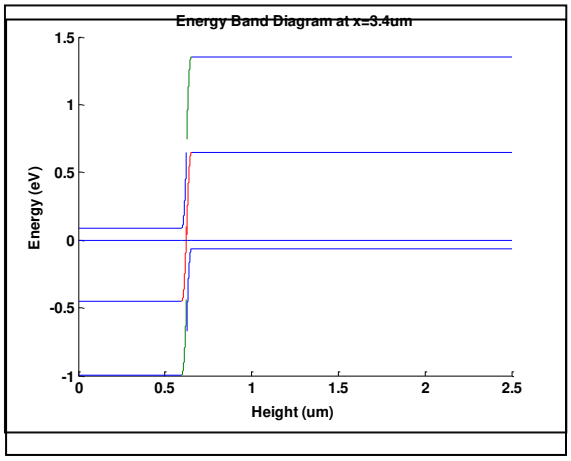


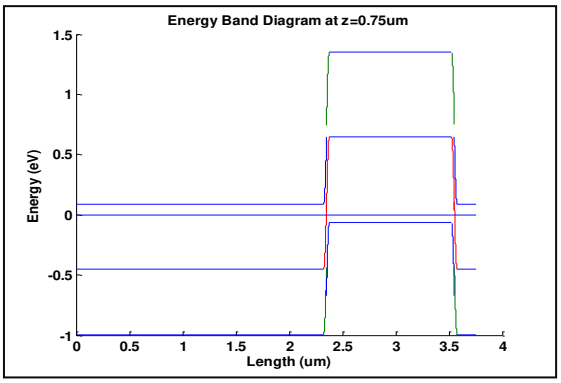
Figure 15. Logic Used For Band Diagram Plot (bdp)



(a)



(b)



(c)

For selected plane x=2um of device structure shown in figure 6, one can infer that regions coming along this plane are region 1 (Si n-type) and region 4 (Si p-type) in sequence, forming a homo-junction. Band structure for this plane is shown in figure 16(a). Second plane chosen is x=3.4um, along which regions coming are 1 and 5, making a hetero-junction, its band structure is shown in figure 16(b). Band structure along two other planes z=0.75um and z=2um are also shown in figure 16(c) and figure 16(d) respectively.

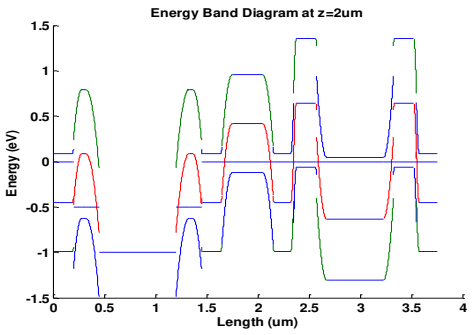


Figure 16. Band Structure of Device at (a) x=2 um; (b) x=3.4 um; (c) z= 0.75 um;(d) z= 2 um

These results are also summarized in table 5 for comparison.

TABLE V. RESULTS

Plane	Region Matrix	Axes (um)
x = 2	[1,4]	[0, 1.25,2.5]
x=3.4	[1,5]	[0, 0.625,2.5]
z= 0.75	[1,5,1]	[0, 2.35,3.55,3.75]
z= 2	[1,2,3,2,1,4,1,5, 6,5,1]	[0,0.2,0.45,1.2,1.45,1.65,2.15,2.3 5,2.575, 3.325,3.55,3.75]

V. CONCLUSION

In this article, we presented a novel algorithm for interactive multi-planar band structure plotting and its implementation using MATLAB®. This work will be of great help to students and device engineers to quickly plot the band diagrams along selected plane, either in horizontal or vertical directions. The proposed algorithm follows Planar Region Matrix (PRM) approach, which first extracts a region matrix on the basis of user provided device specification and plane and then by doing appropriate calculations, plots the band diagram. This algorithm has its own set of conventions, articulately defined above. Various functions of the algorithm are described in various sections using flow charts. The band diagrams were plotted using this algorithm along various planes and are verified for an arbitrary device structure.

REFERENCES

[1] N. Sanjay Rebello, Chandramouli Ravipati, Dean A. Zollman and Lawrence T. Escalada "Computer simulation of p –n junction devices", Am. J. Phys., Vol. 65, No. 8, August 1997.

[2] Richard S. Muller and Theodore I. Kamins, Device Electronics for Integrated Circuits, 3rd Edition**,** Wiley India Edition, 2003.

[3] B. G. Streetman and Sanjay Kumar Banerjee , Solid State Electronic Devices, 6th ed., Pearson Prentice Hall India, 2008.

[4] Peter Y. Yu and Manuel Cardona, Fundamentals of Semiconductors: Physics and Materials Properties, Springer-Verlag Telos; 2nd edition, April 1999

[5] Lev I. Berger, Semiconductor Materials, CRC Press; 1st edition, December, 1996

[6] MATLAB® Package, [Online], https:// www.mathworks.com.

◈◈◈

Clutter Reduction in Ground Penetrating Radar B-Scan Imagery by Gaussian filtering for Improved Landmine Recognition

P. Iswerya

Department Of Electronics and Communication, Indira Gandhi Institute of Technology, GGSIPU, Delhi, India
E-mail: aashi.dec@gmail.com

Abstract - In this paper, a digital image processing technique has been developed to reduce the clutter in GPR B scan images for improved detection of Landmines, using the Image Processing toolbox in MATLAB version 7.0.1. The code developed involves the use of 2-D special filters to create Gaussian low pass and high pass filter. The signals that are unrelated to the target scattering characteristics, i.e., clutter, is effectively reduced and landmine signature areas are highlighted.

Keywords- *Clutter, GPR, Landmines, Image processing, Gaussian filter*

I. INTRODUCTION

War ends when the fighting stops, yet landmines remain dangerous many years after a conflict has ended. The global landmine crisis is creating immense social and economic problems worldwide [1], [2]. According to the International Campaign to Ban Landmines (ICBL), more than 50 countries have produced more than 350 different kinds of AP mines [8]. Removal of landmines, which is also called humanitarian demining, has therefore become one of the major stakes since the beginning of this century. Ground-penetrating radars (GPR) have become key sensors for landmine detection as they are capable of detecting landmines with low metal contents. They have immense detection potential and have been a part of numerous multisensor systems developed these past few years [3]. Ground Penetrating Radar, also known as Georadar, Subsurface Interface Radar, Geoprobing Radar, is a totally non destructive technique to produce a cross section profile of subsurface without any drilling, trenching or ground disturbance.

A. GPR Principles

GPR transmits a pulsed electromagnetic wave from a transmitting antenna located on the ground surface and signals are received by a receiving antenna. The reflection occurs, when the electromagnetic wave encounters any electrically inhomogeneous material. The velocity and reflectivity of the EM wave in soil is characterized by the dielectric constant (permittivity) of the soil. When the dielectric constant of the soil is ε_r the velocity in this material is given by:

$$v = \frac{c}{\sqrt{\varepsilon_r}} \quad (1)$$

where c is speed of light in meters per second.

When electromagnetic wave is incident to a flat boundary of two different materials having the dielectric constant of ε_1 and ε_2, the reflection coefficient of the boundary is defined as

$$\Gamma = \frac{\sqrt{\varepsilon 1} - \sqrt{\varepsilon 2}}{\sqrt{\varepsilon 1} - \sqrt{\varepsilon 2}} \quad (2)$$

The dielectric constant of subsurface material is based on rocks, and soils, which varies in its constituent material itself. However, the dielectric constant of these materials is similar, and the water contained in the material is the most significant for the value of the dielectric constant. Any change of water conditions in the soil and geological formations can cause the electromagnetic reflection [4].

B. GPR Scans

In GPR systems, data can be collected using three scanning geometries, A, B and C scans [5] which are used for inspection of data on the acquisition computer and in laboratory analysis.

A scan depicts the presence of mines by pulses of large amplitude and absence of mines by normal pulses. It is an Amplitude Vs Time graph.

B scan is achieved by collecting a series of A-scan on a horizontal survey line of the earth surface.

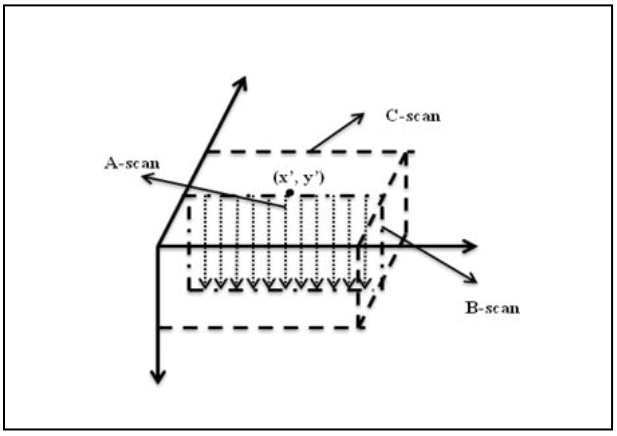


Figure 1. The 3D coordinate system defined on a section of ground

It results from side-by-side 2-D display of a number of traces which are collected at adjacent spatial measurement position during the GPR collection.
B scan is a Time delay Vs Distance graph which exhibits defocused, hyperbolic characteristics. This is mainly due to the finite beam width of the main lobe of the GPR antenna which transmits energy in conical pattern [6].

C scan results from the side-by-side display of a number of GPR sections in a 3-D volume. It is essentially an x, y plane at a selected value of z, or range of values of z [7].

II. CLUTTER

GPRs deliver so-called B scan data, which are, roughly, vertical slice images of the ground. However, due to the high dielectric permittivity contrast at the air–ground interface, a strong response is recorded at an early time by GPRs. This response is the main component of the so-called clutter noise, and it blurs the responses of landmines buried at shallow depths. Small stones and gravel present in the soil also cause clutter. Moreover, the dielectric constant changes with soil moisture. Thus, even if the material of the soil is homogeneous, when moisture is not homogeneous, the electromagnetic wave can be reflected by the soil causing blurring of responses [9]. Coupling between the transmitting and receiving antennae is also a reason for the same. The landmine detection task is hence quite difficult, and a preprocessing step, which aims at reducing the clutter, is often needed. A new and simple clutter removal method based on the design of a two-dimensional Gaussian digital filter, which is adapted to B scan data, is proposed. The designed filter must reduce the clutter on B scan data significantly while protecting the landmine responses.

III. EXISTING DATA PROCESSING TECHNIQUES

Recent developments have shown that significant efforts have been focused towards the extraction of meaningful interpretation from GPR data. Vera, Boriana, Christo [10] calculate of the amplitude and two-way time delay of a signal reflected from each layer of a multi-layered media, simulate useful echo signals, and construct the synthetic range profile. K. Ho and Garder [11], deal with a non-stationary clutter environment and the linear prediction coefficients are computed adaptively. For interpreting buried objects like pipes, tunnels and AP landmines in subsurface images, a variety of techniques applying the Hough transform is used by Aggrawal and Karl [17]. Capineri, Grande, Temple [12] also use the classical Hough transform in order to identify linear segments in the image, representing transitions between layers of different electrical impedances. The authors proposed also a method for extracting hyperbolic signatures of buried objects and hence estimating their position. Delbò, Gamba, Roccato [13] applied a fuzzy clustering approach to identify hyperbolas from GPR images beforehand de-noised. Jeng, Li, Chen and Chien [14] have designed two filters by adopting adaptive algorithms, the optimum 2D median filter, (a 2D median filter with an optimum window size), and the 2D adaptive Wiener filter (a real time optimal filter renovated from the conventional Wiener filter technology) to investigate the advantages of using adaptive filters in processing ultra-shallow seismic and ground-penetrating radar data. Pasolli, Melgani and Donelli [15] propose a novel pattern-recognition system to identify and classify buried objects from ground-penetrating radar (GPR) imagery. The entire process is subdivided into four steps. After a preprocessing step, the GPR image is thresholded to put under light the regions containing potential objects. The third step of the system consists of automatically detecting the objects in the obtained binary image by means of a search of linear/hyperbolic patterns formulated within a genetic optimization framework. Van der Merwe and Gutpa [16] propose an iterative method based on the same idea which takes into account the presence of shallow buried objects and the incoherent component of the clutter, i.e., noise and nondeterministic perturbations. The main drawback of this method is that a reference signature of the buried object is needed.

Al-Nuaimy, Huang, Nakhkash, Fang, Nguyen, Eriksen[18] subdivide the detection process in three main stages: 1) preprocessing step to reduce noise and undesired system effects; 2) image segmentation with an artificial neural network classifier to identify areas potentially containing object reflections; and 3) Hough transform to detect hyperbolic patterns. Gamba and

Lossani [19] implement some preprocessing steps to enhance the signature of buried targets. Then, automatic image interpretation is carried out by a detector based on artificial neural networks.

IV. THE PROPOSED IMAGE PROCESSING METHODOLOGY

In this paper, a software tool developed in Matlab using Image Processing Toolbox (Version 5.0.1) for Ground Penetration Radar is presented. It is intended for basic GPR B scan image processing. The novel system proposed is illustrated in the flowchart given in Fig. 3.

A. Fourier Transform

Frequency domain analysis is carried out using Fourier transform. The Fourier Transform isolates and processes particular image frequencies with a great degree of precision. For images, discrete Fourier transform, abbreviated as DFT is of relevance. Matlab in-built function takes DFT of the matrix image pixels [20].

The two dimensional DFT takes a matrix as input, and returns another matrix, of the same size, as output. If the original matrix values are f(x, y) where x and y are the indices, then the output matrix values are F(u, v). The forward transforms for an MxN matrix, where for notational convenience we assume that the x indices are from 0 to M-1 and the y indices are from to N-1 are:

$$F(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) e^{-p} \quad (3)$$

where $p = 2\pi i(\frac{xu}{M} - \frac{yv}{N})$

The DC coefficient-The value F (0,0) of the DFT is called the DC coefficient. If we put u=v=0 in the definition given in Equation (3) we find that

$$F(0,0) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) e^{0} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \quad (4)$$

That is, this term is equal to the sum of all terms in the original matrix.

For purposes of display, it is convenient to have the DC coefficient in the centre of the matrix. Fig. 2 demonstrates how the matrix is shifted by this method.

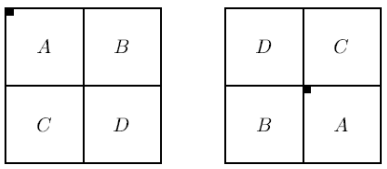


Figure 2. (a) DFT before shift (b) After shifting

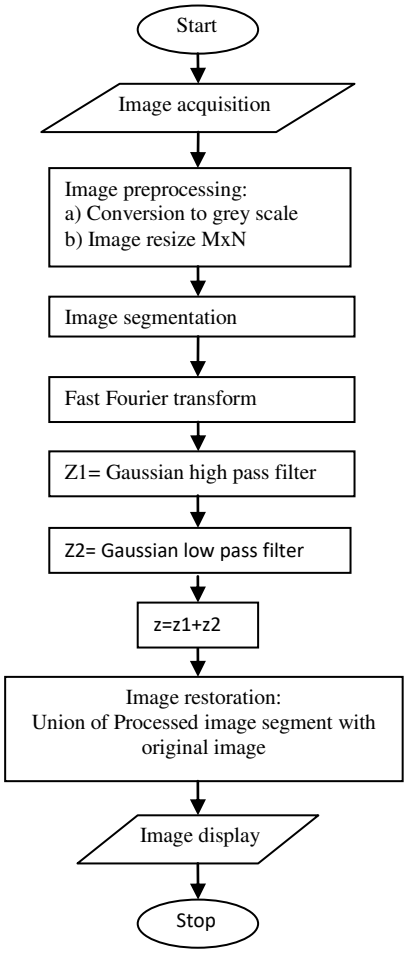


Figure 3. Flowchart for the proposed technique

B. Gaussian Filter

Gaussian filters are a class of low-pass filters, all based on the Gaussian probability distribution function. They may be considered to be the smoothest as compared to the ideal filters and Butterworth filters. A two dimensional Gaussian function is given by:

$$f(x) = e^{-r} \; where \; r = \frac{x^2+y^2}{2\sigma^2} \quad (5)$$

where σ is the standard deviation.

A Matlab command produces a discrete version of this function with parameters being filter size, F_S and standard deviation, S_D.

F_S (which is also optional), is a vector specifying the number of rows and columns in the generated filter f(x), or a scalar indicating equal size for height and width, the default value being 3x3. S_D marks the thinness of the Gaussian filter, which if not given, defaults to 0.5. A large value of standard deviation produces a flatter curve, and a small value leads to a "pointier" curve.

Gaussian filters have a blurring effect which looks very similar to that produced by neighborhood averaging.

C. Algorithm for clutter reduction

The entire algorithm can be divided into five steps; (1) Image acquisition (2) Image preprocessing (3) Image segmentation (4) Gaussian filtering (5) amalgamating the image with the input image.

The paradigm for the same can be elaborated as follows:

Once the GPR B scan image is read, it is resized to a standard size of 256x256. The aim is to do away with clutter, which is by and large present in the upper layers of the soil. Thus the upper portion of the image, about 30%, is separated for fast filtering purpose.

In the second phase, Fourier transform is applied to the image and the DC coefficient is shifted as discussed in equation (4.2). Gaussian high pass and low pass are then generated. In order to design the filter, the parameters F_S, S_D and N must be determined, where N is the size of the segmented image. Smaller the value of N, higher is the computational speed. These filters are element-wise multiplied to the image transform, and inverse Fourier transform is obtained for both.

Combining the above obtained results, Gaussian band pass filter is achieved. The process ends by merging the above processed image with original image.

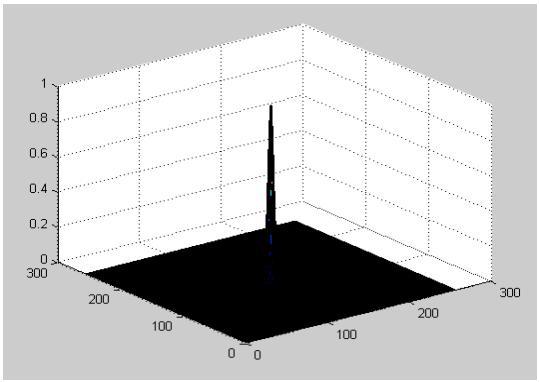
V. TEST RESULTS

In this section some results achieved using the above presented software tool are shown.

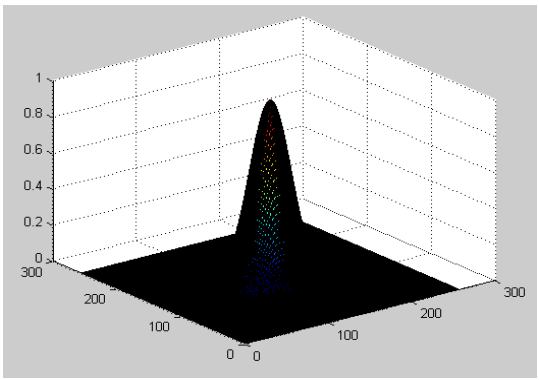
The surface plots of the designed Gaussian filter for varying σ is shown in Fig. 4. G2 is the Gaussian low pass filter with F_S = 265 and S_D = 3. G1 and H1 the Gaussian low pass and high pass filter with F_S = 265 and S_D = 20.

The MACADAM (Multisensor Acquisition Campaign for Analysis and Data Fusion of Antipersonnel Mines) campaign real B scan data are used to test the system described in previous section, as demonstrated in Fig 5.

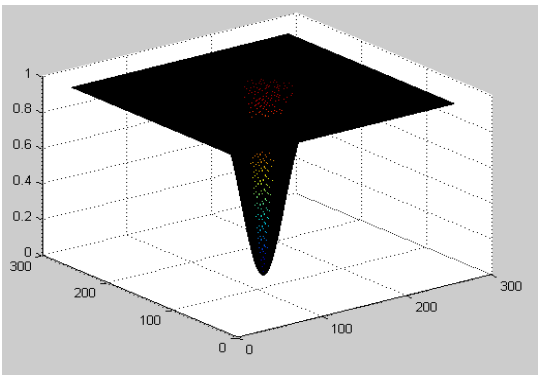
B scan in fig. 5(a) shows three clutter bands above the landmine which are removed to a great extent. Also, the landmine signatures have been highlighted for improved detection. Value of N is taken to be 70.



(a)



(b)



(c)

Figure 4. Surface plot (a) G2 (b) G1 (c) H1 (=1- G1)

The proposed design is now tested on the B scan shown in Fig. 5(c). Two VSMK2 landmine responses have been recorded, but they are hardly visible because their responses overlap with the one from the clutter. As shown in Fig. 5(d), after digital filtering with N = 70, the three horizontal bands have been filtered, but there are some residues.

Consider the B scan displayed in Fig. 5(e). In this image, two landmine responses of MAUS1 type (metallic content) have been recorded. The ground is

made of clay covered by grass. The result is depicted in Fig. 5(f) for N = 90. The three bands representing the clutter are well filtered, and the areas with landmine signatures have been highlighted with respect to the neighboring portions.

Hence, the clutter is reduced by the blurring action of Gaussian filter. The areas with landmine signatures have been darkened with respect to the neighboring portions as well as the signatures, which can be clearly seen in Fig. 5 (b), (d) and (f). The background of processed image is darker than the original image due to processing technique used. Nevertheless, landmines can be identified. Thus the tool improves landmine detection by minimizing clutter in B scan GPR images by means of Gaussian filter.

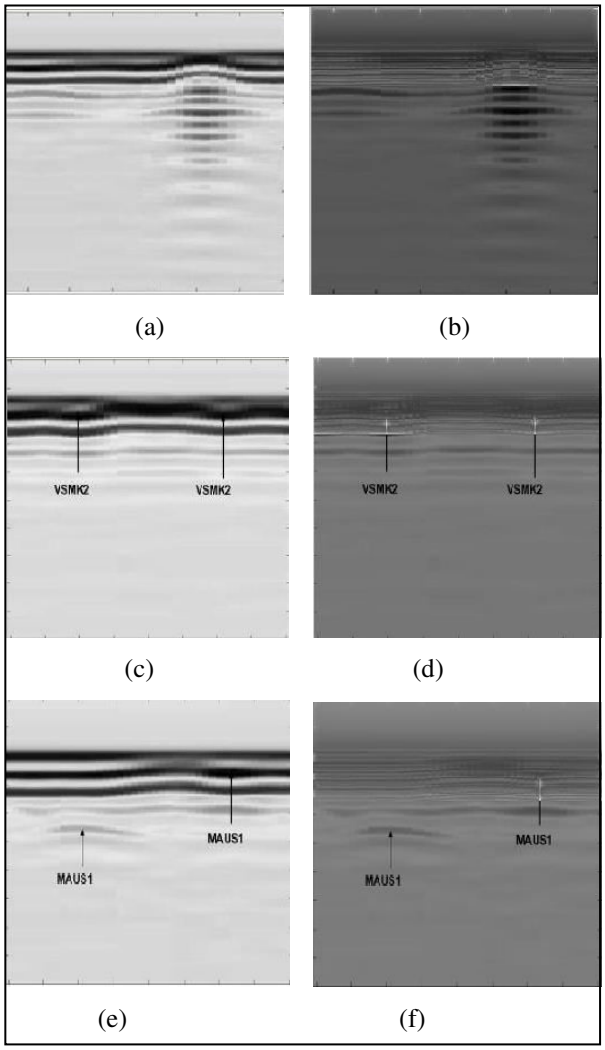


Figure 5. (a) Original Test image1 (b) Processed Test image1 (c) Original Test image2 (B scan recorded above an agricultural soil.) (d) Processed Test image2 (e) Original Test image3 (B scan recorded above an agricultural soil.) (f) Processed Test image3

VI. CONCLUSION

This work presents a new digital image processing technique to reduce clutter in GPR data. A Gaussian filter is adapted that reduces the clutter due to subsurface as well as moisture inhomogeneities. The different test results show the efficiency of such a filter while protecting the landmine signatures. The design works well for the test images when appropriate filter parameters such as standard deviation and filter size are selected. Clutter can be removed up to varied depths, thereby making the code very flexible for use. Moreover, this code works well irrespective of the fact that the clutter bands are horizontal or not. The implementation of such a filter is simple and its computational cost is low. Thus, the proposed technique has potential for real world applications.

ACKNOWLEDGMENT

The author is highly grateful to Mr. Ashish Gautam, scientist 'B', DTRL lab, DRDO for his valuable guidance. She would like to extend her sincere thanks to him for the insightful discussions and his constant interest, encouragement and support throughout the development of the research.

REFERENCES

[1] K. Tsipis, "Report on the landmine brainstorming workshop," Tech. Rep., Sci. Technol. Int. Security Prog., Mass. Inst. Technol., Cambridge, 1996.

[2] J. Hurst, "Ten million tragedies, one step at a time," Bull. At. Scientists, pp. 362–367, July/Aug. 1993.

[3] Delphine Potin, Emmanuel Duflos, Philippe Vanheeghe," Landmines Ground Penetrating Radar Signal Enhancement by Digital Filtering", IEEE Transactions on Geoscience & Remote Sensing , vol. 44, no. 9, September 2006.

[4], Motoyuki Sato," Principles of Mine Detection by Ground-penetrating Radar,Tohoku University, Sendai 980-8576, Japan

[5], C. Ozdemir, S. Demirci, and E. Yigit,"Practical algorithms to focus B-Scan GPR images: Theory and application to real data", Progress in Electromagnetics Research B, Vol. 6, 109–122, 2008

[7] Daniels DJ, Ground Penetrating Radar (2nd Edition) Knoval (Institution of Engineering and Technology) pp. 1-4, (2004).

[8] Galal Nadim,"Clutter Reduction and Detection of Landmine Objects in Ground Penetrating Radar Data Using Likelihood Method", ISCCSP 2008, Malta, 12-14 March 2008.

[10] Vera Behar, Boriana Vassileva, Christo Kabakchiev," A Software Tool for GPR Data Simulation and Basic Processing", Cybernetics and Information Technologies, Volume 8, No 4, Sofia,2008.

[11] K. Ho and P. D. Garder, "A linear prediction landmine detection algorithm for hand held ground penetrating radar," IEEE Transactions on Geoscience & Remote Sensing , vol. 40, no. 6, pp. 1374–1384, Jun. 2002.

[12] L. Capineri, P. Grande, J. A. G. Temple, "Advanced image-processing technique for real-time interpretation of ground-penetrating radar images", International Journal of Imaging Systems and Technology, vol. 9, pp. 51-59, 1998.

[13] S. Delbò, P. Gamba, D. Roccato, "A fuzzy shell clustering approach to recognize hyperbolic signatures in subsurface radar images", IEEE Transactions on Geoscience & Remote Sensing , vol. 38, pp. 1447-1451, 2000.

[14] Y. Jeng, Y. Li, C. Chen and H. Chien, Adaptive filtering of random noise in near-surface seismic and ground-penetrating radar data, Journal of Applied Geophysics, Vol.68, No.1, May 2009, pp36-46.

[15] E. Pasolli, F. Melgani and M. Donelli, Automatic Analysis of GPR Images: A Pattern-Recognition Approach, IEEE Transactions on Geoscience & Remote Sensing, Vol.47, No.7, Jul 2009, pp. 2206-2217.

[16] A. Van der Merwe and I. J. Gutpa, "A novel signal processing technique for clutter reduction in GPR measurements of small, shallow land mines," IEEE Transactions on Geoscience & Remote Sensing , vol. 38, no. 6, pp. 2627–2637, Nov. 2000.

[17] Aggrawal N. and Karl W.C., "Line detection in images through regularized Hough transform", IEEE Transaction on Image Processing vol. 15, pp 582-590,2006.

[18] W. Al-Nuaimy, Y. Huang, M. Nakhkash, M.T.C. Fang, V.T. Nguyen, A. Eriksen, "Automatic detection of buried utilities and solid objects with GPR using neural networks and pattern recognition", Journal of Applied Geophysics, vol. 43, pp. 157-165, 2000.

[19] P. Gamba, S. Lossani, "Neural detection of pipe signatures in ground penetrating radar images", IEEE Transactions on Geoscience & Remote Sensing , vol. 38, pp. 790-797, 2000.

[20] Rafael Gonzalez and Richard E.Woods. Digital Image Processing, Addison-Wesley, 2nd edition, 2002

◈◈◈

Error Performance of Turbo Codes in AWGN Channels

P. Surendra Kumar, B.Chandra Mohan & T.Krishna Chaitanya

Department of ECE, Bapatla Engineering College, Bapatla, Guntur (Dist.), A.P., India.
E-mail : mailme_surendranitk@yahoo.co.in, chandrabhuma@yahoo.co.in, tkrishna479@ gmail.com

Abstract - This paper presents an error-control code and the software implementation of two decoders, namely the Maximum a Posteriori (MAP) and Soft- Output Viterbi Algorithm (SOVA) decoders. A comparison of perform is made against that of the classical Viterbi decoder. They achieve near-Shannon-limit error correction performance with relatively simple component codes and large interleavers. Error-correcting capability of the code can be readily investigated by modifying the software routines. It is also shown that the MAP algorithm gives a better error performance than the SOVA decoder under similar conditions.

Keywords- *MAP, SOVA, Interleaver, Deinterleaver, Turbo decoding.*

I. INTRODUCTION

A software-based project in error-control coding combines theory and design with computer simulation. It helps to grasp the principles of communication systems, the need to add controlled redundancy via a channel encoder and to exploit this redundancy at the receiver side, in order to recover the transmitted information through a channel decoder. Further, it helps to understand Shannon's theorem and the concept of the Shannon limit. This paper presents a project which consists of a study of turbo codes as an error-control code and the software implementation of two different decoders, namely the Maximum a Posteriori (MAP) and Soft- Output Viterbi Algorithm (SOVA) decoders. A comparison of their performances is made against that of the classical Viterbi decoder.

Turbo codes were introduced in 1993 and are perhaps the most exciting and potentially important development in coding theory in recent years. They achieve near-Shannon-limit error correction performance with relatively simple component codes and large interleavers. They can be constructed by concatenating at least two component codes in a parallel fashion, separated by an interleaver. One feature of turbo codes is that the constituent codes need not be complex. Simple (2,1,4) convolutional codes can achieve very good results. In order for a concatenated scheme such as a turbo code to work properly, the decoding algorithm must effect an exchange of soft information between component decoders. The concept behind turbo decoding is to pass soft information from the output of one decoder to the input of the succeeding one, and to iterate this process several times to produce better decisions. Turbo codes are still in the process of standardization but future applications will include mobile communication systems, deep space communications, telemetry and multimedia.

II. TURBO CODE ENCODER

A turbo encoder is the parallel concatenation of recursive systematic convolutional (RSC) codes, separated by an interleaver, as shown in Fig. 1. The data flow d_k goes into the first elementary RSC encoder, and after interleaving, it feeds a second elementary RSC encoder. The input stream is also systematically transmitted as X_k, and the redundancies produced by encoders 1 and 2 are transmitted as Y_{1k} and Y_{2k}. For turbo codes, the main reason of using RSC encoders as constituent encoders instead of the traditional non-recursive nonsystematic convolutional codes, is to use their recursive nature and not the fact that they are systematic.

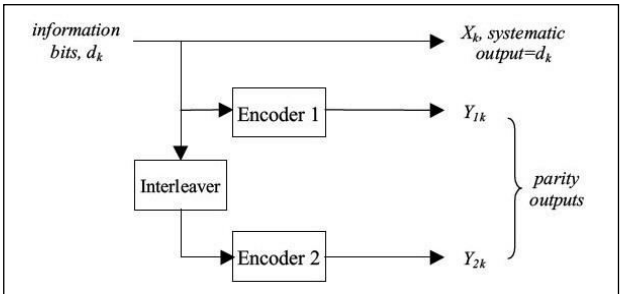


Fig. 1 *A fundamental turbo encoder.*

The global rate of the turbo encoder in Fig. 1 is one-third. To achieve a higher rate, the parity outputs can be punctured. In order to obtain a rate half encoder, the parity bits at the encoder outputs are selected alternately, i.e., one bit from encoder 1 and then one bit from encoder 2, as illustrated in Fig. 2. The interleaver is an important design parameter in a turbo code. It takes a particular stream at its input and produces a different sequence as output. Its main purpose at the encoder side is to increase the free distance of the turbo code, hence improving its error-correction performance.

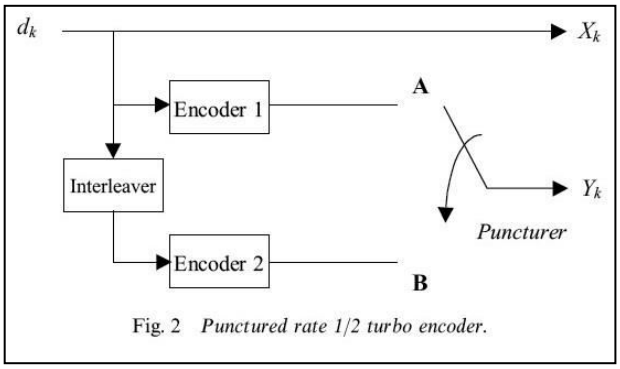


Fig. 2 *Punctured rate 1/2 turbo encoder.*

There are different types of interleavers, such as the block, pseudo-random, simile, and odd-even interleavers. They differ in the way they shuffle the input symbols. As an example, the block interleaver is explained below. A sequence of L bits is written into an N×M matrix row by row starting from the first row of the matrix. Block interleaving then consists of reading the matrix elements column by column starting from the first one. The resulting sequence is written to an array of length L as shown in Fig. 3.

0	1	2	3
4	5	6	7
8	9	10	11
12	13	14	15

Input Sequence 0,1,2,3,…,12,13,14,15

Output Sequence 0,4,8,12,…,3, 7, 11,15

Fig. 3 *Block interleaving.*

III. TURBO DECODING

A turbo code is far too complex to decode with a single decoder. Instead, each convolutional code in the turbo code is decoded separately with soft information being passed from one decoder to the next. The decoding scheme of a turbo code is shown in Fig. 4. The above decoder consists of two serially interconnected soft-in soft-out (SISO) decoders, which can be SOVA or MAP decoders. x_k and y_k are the channel outputs with x_k corresponding to the systematic encoder output, and y_k is a multiplexed stream of the two punctured encoder outputs. Hence, a demultiplexer is necessary at the receiver. z_k is called the a priori information and is equal to zero for the first iteration. The decoder soft output, L (d_k), also called log-likelihood ratio (LLR), can be separated into three components:

$$L(d_k) = L_{sys} + L_{apr} + L_{ext} \text{ ----------(1)}$$

d_k denotes the actual hard decision of the decoder at step k. L_{ext} is called the extrinsic information. It is a function of the redundant information introduced by the encoder and has the same sign as d_k.7 (L_{sys}+L_{apr}) constitutes the intrinsic information. L_{sys} is the LLR for the channel output and is the received systematic input to the decoder, scaled by the channel reliability. L_{apr} is equal to the extrinsic information produced by the previous decoder.

The intrinsic information is subtracted from the soft decoder output (LLR1, LLR2). The resulting output is the extrinsic information, which is passed on to the next decoder. This process is repeated until a desired performance is attained after a number of iterations.

It is the decoding method that gives turbo codes their name, since the feedback action of the decoder is reminiscent of a turbo-charged engine.

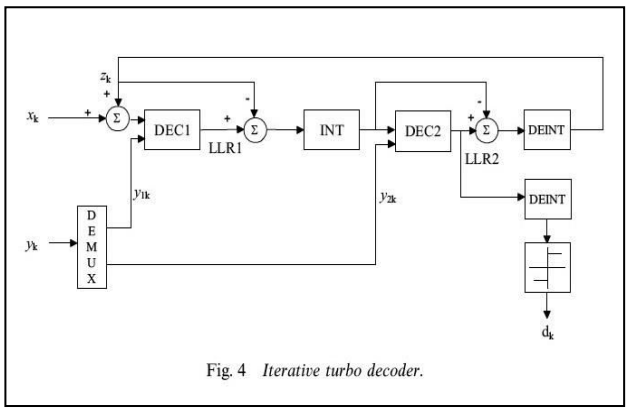


Fig. 4 *Iterative turbo decoder.*

IV. PRINCIPLE OF THE SOVA

The SOVA is a modified Viterbi algorithm which produces soft outputs associated with the decoded bit sequence. These modifications include a modified metric computation and a reliability value update along the maximum likelihood (ML) path.

Let m denote a trellis branch into a node and $M^{(m)}{}_k$ denote the accumulated metric at time k for branch m. Also, let $u_k^{(m)}$ be the systematic encoder output (in bipolar voltage form) for path m and $x_k^{(m)}$ be the corresponding parity output (Fig. 5).

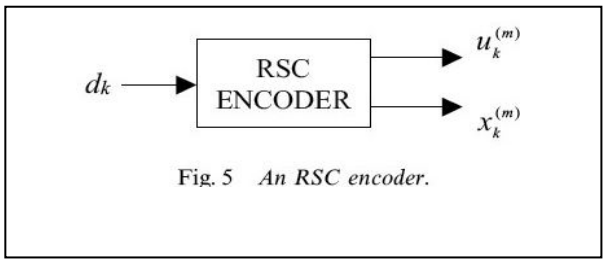


Fig. 5 *An RSC encoder.*

If $y_{k,1}$ and $y_{k,2}$ are the channel outputs corresponding to the systematic and parity outputs of the RSC encoder, then the metric used for the SOVA algorithm becomes4

$$M_k{}^m = M_{k-1}{}^m + u_k{}^m L_c y_{k,1} + x_k{}^m L_c y_{k,2} \quad \text{-----(2)}$$

Referring to the trellis section in Fig. 6, there are two branches merging into the central node (i.e. the node at time k). Therefore, m=1, 2. In the first step, the metric associated with each trellis node is computed by performing a forward sweep through the trellis. In addition, the metric difference (which represents the reliability value) between the paths merging in any node is calculated and stored. Then, the ML path is found. Finally, the reliability update procedure is performed as follows:

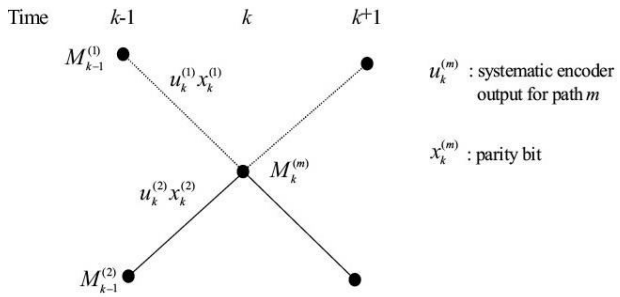


Fig. 6 *Trellis section illustrating SOVA metric computation.*

V. PRINCIPLE OF THE MAP ALGORITHM

The MAP algorithm proceeds by estimating the log-likelihood ratios (LLR) of the decoded bits, based on the received symbols. The MAP decoder produces estimates, d_k, of the information bits based on the comparison of the LLR, $L(d_k)$, to a threshold. For equiprobable binary input values over an AWGN channel,

d_k={1 if $L(d_k) \geq 0$ ----------(3)

0 if $L(d_k) < 0$

Defining $\alpha_k^i(m)$ and $\beta_k^i(m)$ as forward and backward state metrics respectively, where m is a state on the trellis diagram, the LLR is given by

$$L(d_k) = \log \frac{\sum_m \alpha_k{}^1(m)\beta_k{}^1(m)}{\sum_m \alpha_k{}^0(m)\beta_k{}^0(m)} \quad \text{---------(4)}$$

$\alpha_k^i(m)$ ***Represents a state likelihood at time k for input bit i, based on a forward*** sweep through the trellis, and $\beta_k^i(m)$ also represents a state likelihood but it is based on a backward sweep through the trellis. The expression for $\alpha_k^i(m)$ is

$$\alpha_k{}^i(m) = \delta_i(R_k, m)\sum_{j=0}^{1} \alpha_{k-1}^j\left[S_b^j(m)\right] \quad \textbf{----------(5)}$$

where $S^j{}_b(m)$ is the state going backwards in time which leads to state m with input j. $\delta_i(R_k, m)$ is the branch metric and depends on the channel. It is given by

$$\delta_i(R_k, m) = \exp\left\{\frac{2}{\sigma^2}\left[x_k i + y_k Y_k^i(m)\right]\right. \quad \textbf{----------(6)}$$

where $Y^i{}_k$ (m) is the encoder output with input bit i and encoder state m. Also,

$$\beta_k^i(m) = \sum_{j=0}^{1} \beta_{k+1}^j\left[S_f^i(m)\right]\delta_j\left[R_{k+1}, S_f^i(m)\right] \quad \textbf{-----(7)}$$

where S_i^f f(m) is the state going forward in time starting from state m with input bit i. Using (5) and (7), the state metrics $\alpha_k^i(m)$ ***and*** $\beta_k^i(m)$ can be computed recursively. On a trellis, $\delta_i(R_k, m)$ corresponds to the branch of the transitions from time k to k+1, with an initial encoder state m and an information bit i

The MAP algorithm is as follows:

Step 1

Starting at time k=0, the branch metrics $\delta_i(R_k, m)$ are calculated and stored for all received symbols (x_k, y_k) using (6).

Step 2

Since the encoder starts in state 00, $\alpha_0^i(m)$ is initialized to $\delta_i(R_0, 00)$ for i=0, 1 and $\alpha_0^i(00)$ for all m $m \neq 00$. $\alpha_k^i(m)$ is computed for each k from 1 to N−1 using (5).

Step 3

Since the trellis is terminated in the zero state, $\beta^i_{N-1}[S^i_b(00)]$ is initialized to 1 for i=0, 1 and $\beta^i_{N-1}(m) = 0$ for all other $m \neq 00$ $\beta^i_k(m)$ is computed and stored for each k from N−2 to 0 using (7).

Step 4 Finally, using (4), L (d_k) is computed for k=0 to N−1 and the estimated bit d_k is obtained using (3).

VI. DESIGN AND SIMULATION

The aim of this project is to to have hands-on experience of:

(a) Generation of a pseudorandom sequence,

(b) Turbo encoding of bits using the parallel concatenated RSC encoders,

(c) Generation and addition of AWGN to the encoded bits,

(d) Iterative decoding of the corrupted information using SOVA and MAP algorithms,

(e) Computation of the BER at different levels of signal-to-noise ratio per bit, and

(f) Evaluation of the effect of using different frame sizes.

In working through the project, the effect of channel impairments on the data transmitted and how the errors are reduced through the use of error-control codes. In addition, compares the performance of these codes using the Viterbi, SOVA and MAP decoding approaches.

Moreover, the effect of increasing the frame size is made clear.

The project is divided into three parts:

(a) The generation of AWGN and evaluation of its effect on the performance during data transmission using hard decision Viterbi decoding,

(b) Study of the effect of employing a convolutional code during data transmission in a noisy channel, and

(c) Study of the performance improvement produced by employing turbo codes during data transmission in a noisy channel.

The three schemes to be implemented are illustrated in Fig.7.

VII. TYPICAL RESULTS AND EVALUATION

The variation of the bit error rate with the signal-to-noise ratio per bit has been measured for the two decoding methods, using different frame sizes. Using bipolar (antipodal) signaling scheme, 100000 channel code symbols are transmitted over a discrete noisy channel in each test. For all cases, perfect bit and frame synchronization is assumed. In each test, the average energy per information bit is fixed, and the variance of the AWGN is adjusted for a range of average bit errors.

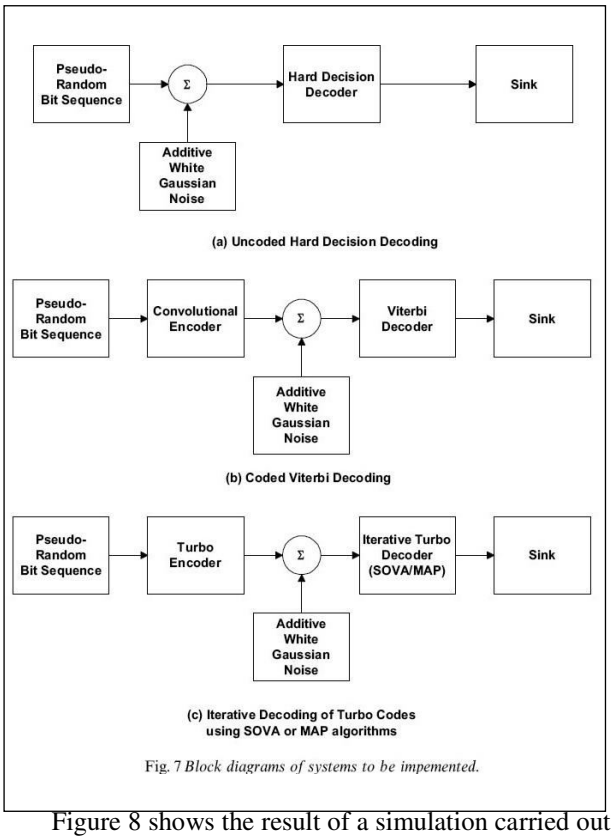


Fig. 7 *Block diagrams of systems to be impemented.*

Figure 8 shows the result of a simulation carried out to compare the performance of the SOVA and the MAP algorithm using a 4-state convolutional encoder and a frame size of 64 bits. From Fig.8, it is seen that the MAP algorithm gives the best performance in terms of BER performance, followed by the SOVA and the Viterbi algorithm. At a BER of 10−4, the MAP algorithm gives a coding gain of 0.4 dB over the Viterbi algorithm while the performance of the SOVA is similar to that of the Viterbi algorithm since the only difference between these two algorithms lies in the reliability update in the SOVA. Iterative decoding was performed on a rate half 16-state encoder using a frame size of 192 bits for both the SOVA and MAP algorithms. The results are shown in Figs. 9 and 10. It can be seen that as the number of iteration increases, the error performance for both SOVA and MAP decoders is improved. This is due to the sharing of the extrinsic information generated by one decoder with the next one. Hence students understand that this process of sharing of information enables the decoder to make a more accurate decision. It can also be observed that the MAP algorithm performs better than the SOVA decoder. Figures 11 and 12 show the performance of the SOVA and MAP algorithms for

a frame size of 2048 bits, respectively. The simulations were carried out on a 16-state punctured rate half turbo code. It can be noted that as the frame size increases, the error performance of both SOVA and MAP decoders improves. After 5 iterations, the coding gain of MAP decoding over SOVA decoding is about 2 dB when a frame size of 2048 bits is used.

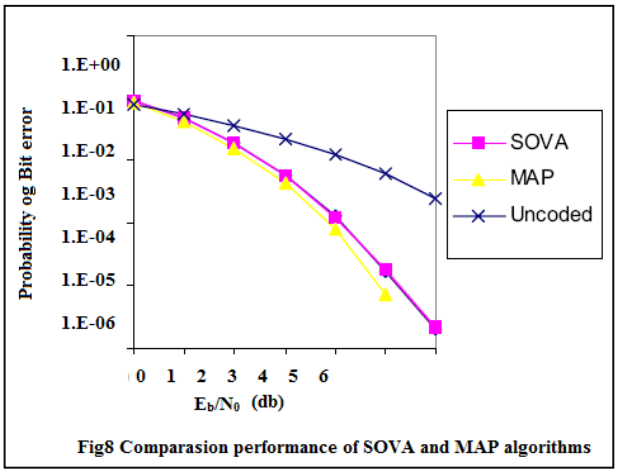


Fig8 Comparasion performance of SOVA and MAP algorithms

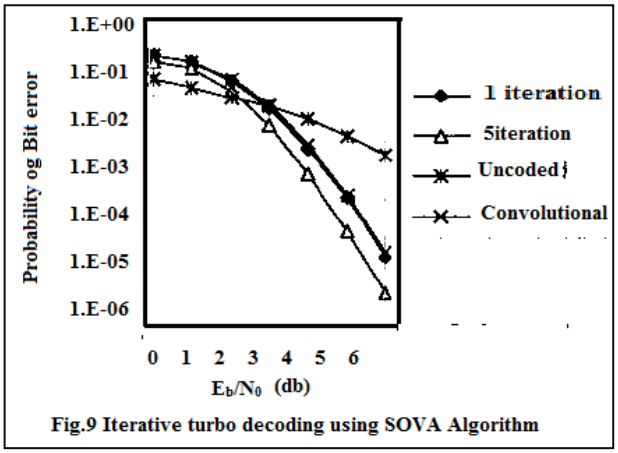


Fig.9 Iterative turbo decoding using SOVA Algorithm

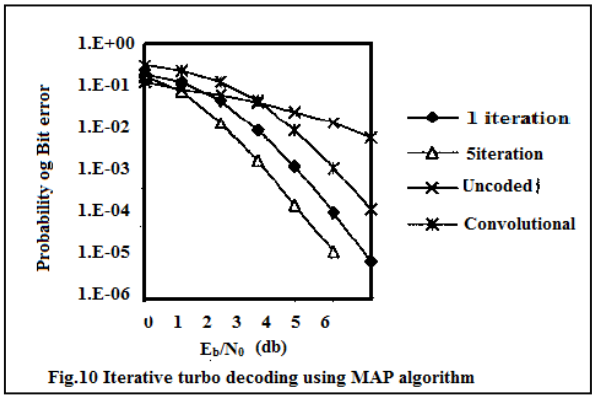


Fig.10 Iterative turbo decoding using MAP algorithm

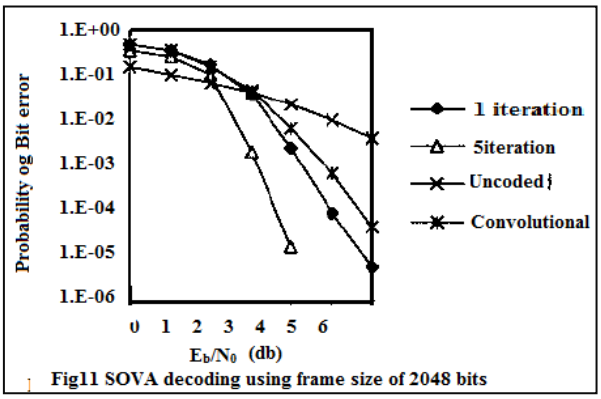


Fig11 SOVA decoding using frame size of 2048 bits

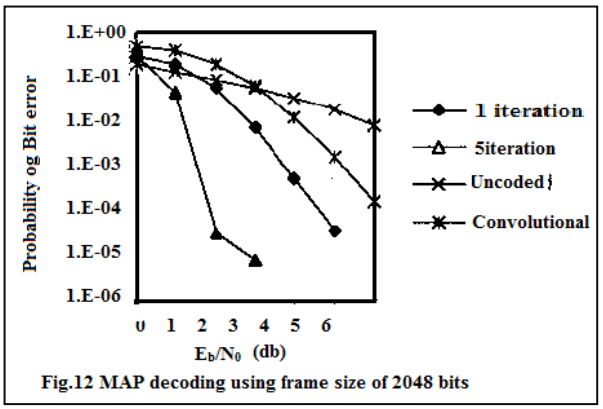


Fig.12 MAP decoding using frame size of 2048 bits

VIII. CONCLUSIONS

A design and simulation project that allows students to acquire hands-on experience on the development and analysis of a turbo codec in AWGN channels has been presented in this paper. Being entirely software-based, such a project provides a flexible means for exploring a wide range of schemes around the designated system. Thus, the effect of varying the frame size on the error-correcting capability of the code can be readily investigated by modifying the software routines. It is also shown that the MAP algorithm gives a better error performance than the SOVA decoder under similar conditions.

REFERENCES

[1] Closed Chain Error Correction Technique for Turbo Product Codes. Newcastle University, School of Electrical, Electronic and Computer Eng., Newcastle upon Tyne, U.K. Senior Member, IEEE, and B. S. Sharif CSNDSP 2010.

[2] Soft Incremental Redundancy for Distributed Turbo Product Codes, IEEE Communications Society subject matter experts for publication in the IEEE CCNC 2010 proceedings.

[3] A. AlDweik and B. Sharif, "Nonsequential decoding algorithm for hard iterative turbo

product codes," IEEE Trans.Commun. vol. 57, pp. 15451549, June 2009.

[4] T. F. Wong and J. M. Shea, "Multi-dimensional parity-check codes for bursty channels," in Proc. 2001 IEEE Int. Symp. Inform. Theory, Washington, D.C., June 2001.

[5] J. M. Shea, "Improving the performance of turbo codes through concatenation with rectangular parity check codes," in Proc. 2001 IEEE Int. Symp. Inform. Theory, Washington, D.C., June 2001.

[6] J. M. Shea and T. F. Wong, "Concatenated codes based on multidimensional parity check codes and turbo codes," in Proc. 2001 IEEE MILCOM, Oct. 2001.

◈◈◈

Signal Processing and Fault Detection in Induction Motors

Shrawani Badoniya, Sulochana Wadhwani & A.K.Wadhwani

Abstract - The successful development of an effectual fault detection scheme is obtained by developing an approach or methodology that could effectively be employed to characterize or quantify the normal operating conditions of a motor and is subsequently competent in identifying deviatory behavior as a fault in an unblemished manner. Fault or failure detection is an extensive discipline that incorporates two established approaches for the fault detection, namely, model-free feature extraction methods and model referenced methods. Model free methods are also known as non-deterministic feature extraction methods whereas model reference methods are deterministic methods. This paper discusses the model free approach based on analysis if stator currents of three phase induction motor for fault detection and identification. The proposed method uses Parks transformation approach and gives correct classification of bearing faults.

Keyword- *Induction Motor, Bearing, Stator Current, Park Transformation, Park Pattern..*

I. INTRODUCTION

In many situations, vibration methods are effective in detecting the presence of faults in Induction motor. However vibration transducer is generally placed on only most expensive and critical motors where the cost of condition monitoring can be justified. Also the vibration centre are limited in their ability detect electrical faults. However stator current monitoring can be implemented inexpensively on most of the machines by utilizing current transducer which is placed on motor control center or switchgear. Furthermore the use of current sensor is convenient for monitoring large number of motor remotely from one location. Basically there are two methods which can be adapted for fault analysis and identification as given in next section.

II. MODEL FREE METHODS

Model free methods are also known as non-deterministic feature extraction methods. Three broad categories for model-free methods are: Data driven methods or Signal based methods such as Spectral Analysis, Hardware Redundant systems, and Knowledge based methods such as Knowledge based Expert Systems etc.

Data-driven methods (also recognized as signal-based methods) make use of the sensor data acquired for the duration of normal operating conditions and also during specific faults, to bring to fruition the fault indicators that could effectively and conveniently be employed for detecting and identifying fault(s). These methods make use of inputs such as mechanical vibration, acoustic noise, ultrasonic, current, or voltage and various other sensor signals to detect and diagnose faults. These methods are extensively employed to monitor the health conditions of rotating machinery like induction motors, generators, bearings, gears, pumps etc. By keeping an eye on the trend of the amplitude and frequency distribution of measured signals and weighing them against the characteristics of the healthy system's signals, the health condition of the machine could be assessed.

Hardware redundant systems entail extra hardware that is exclusively dedicated to detect faults. Limit switches, level switches, and redundant sensors are few components that are invariably employed for this particular purpose.

Knowledge-based techniques encompass exceptional methodologies and procedures that could be employed to acquire the momentous and fruitful information that otherwise remain concealed in model-based methods.

The use of model free methods for fault detection and analysis is dictated by two important factors: First, model free methods require a prior knowledge of system signal characteristics. Second, the guiding relationship between the system and system signals would inevitable change with changes in the system operating point. This is particularly important in case of dynamic systems, where the operating point is destined to change very fast during the system operation.

III. MODEL-REFERENCED METHODS

Model-Referenced fault detection is rooted in the concept of analytical redundancy. These methods are deterministic methods and are based on generating a hardware or/and analytical model of the system that is being explored for fault(s). Model-based methods compare analytically generated signals with physically generated signals to decide when a fault has occurred. The system model, usually in the form of a set of equations or ordinary differential equations (ODEs), that provides the analytically generated signals. These techniques entail as many variables and system parameters as possible so as to construct a detailed and precise mathematical model of the system under observation. The use of neural networks imparts the competence for the acceptance of model-based fault diagnosis with conviction

IV. FAULT DETECTION

On the whole, in case of rotating electrical machinery, information present in electrical and mechanical signals can be exploited for fault detection. Typically, vibration measurements have been employed, but recently use of electrical signals, particularly stator current signals has attracted significant focus. Several other techniques such as acoustic noise measurement, torque profile analysis, temperature analysis and magnetic field analysis have been conventionally used for detecting induction motor faults.

The vibration monitoring is one of the time-honored techniques for detection of electrical and mechanical faults in electrical rotating machines in general and induction motors in particular. Usually, the measured vibration signal is the vibration velocity acquired at the bearing housing of the machine. In many situations, vibration The stator current monitoring or motor current signature analysis (MCSA) is relatively a recent technique that is fast gaining importance due to its non-invasive nature, cost-effectiveness, preciseness in analysis, easy and efficient signal processing, and convenient installation. The stator current monitoring and analysis can provide as much information as the vibration monitoring provides. The measured vibration and associated current harmonics are closely correlated. Electrical faults such as broken rotor bars/end rings and unbalanced input supply in induction motors are divulged as imbalance and disturbance in the stator currents. Frequently occurring mechanical faults such as bearing damage, static and dynamic eccentricity, instigate variations in air-gap permeance and air-gap flux density. These changes in the air gap flux are subsequently reflected in stator line currents. It is mostly necessary to confide in time-series data generated from sensors due to unavailability of accurate and trustworthy modelling of the induction machines. The required task of fault detection thus converges to estimation of a fault featuring parameter of interest, given the model of the signal with unknown parameters. Many tools and techniques based on profile and trend analysis of various machine parameters have been used for the condition and fault monitoring in induction Motors. Motor Current Signature Analysis (MCSA) is one such technique that deals with identifying these characteristic frequencies in motor stator currents. Figure 1 shows signal processing approach to fault diagnosis.

V. MOTOR CURRENT SIGNATURE ANALYSIS (MCSA)

This is a non-invasive and non-destructive technique in which the motor supply current (stator current) is chronicled and its analysis (mostly frequency domain spectrum analysis) is used to detect and diagnose the presence of faults in rotating electrical motors and/or drives. MCSA can be performed under both transient and steady state conditions and can be used to diagnose electro-mechanical faults. The work carried out in this research is based on acquiring the motor stator current data through data acquisition system and then analyze it through various time domain, frequency domain and time-scale analysis techniques to detect various induction motor faults.

Both load and fault sources result in disturbances to the air gap flux waveform which consecutively can bring about current components in the motor stator winding. These specific and idiosyncratic components can be perceived by monitoring the current in a supply cable line conductor through a current transducer. The distinctive fault features are then used to provide a basis for fault diagnosis.

As has been mentioned in the literature, the machine faults can be distinguished and analyzed through the quantitative representation of time domain features. This section explores the utility and relevance of a time domain analysis technique namely the Park's vector. It begins with providing the basic details about the Park's vector technique followed by the description and discussion on the related results in detecting various induction motor faults. The MATLAB is employed to obtain Park's vector from the saved stator current three-phase data.

VI. PARK'S VECTOR OF STATOR CURRENT

The stator and rotor faults are ultimately reflected in the stator current and in most of the cases cause unbalanced phase currents. This unbalance can easily be marked through Park's vector analysis.

The three-phase quantities (e.g. Voltage, Current etc.) associated with three-phase induction motors are

expressed in terms of complex space vectors. This complex three-phase space vector model of three-phase induction motor is valid for any instantaneous variations in current and voltage and holds good for both steady-state and transient analysis. The space vectors can be adequately described using only two orthogonal axes. This reduces the three-phase induction machine to an equivalent two-phase machine model. The two-phase model greatly simplifies the solution of otherwise complex current and voltage equations in three-phase motor model.

VII. PARK'S TRANSFORMATION

The Park's transformation is used at replacing the variables such as current, voltage and flux associated with the stator windings of rotating with variables linked with a rotating frame that rotates with the rotor. At a given instant of time, the d axis of the rotor is assumed to be inclined at angle θ radians with respect to a fixed reference frame. The q-axis of the rotor is assumed to be orthogonal to the rotor d-axis. The relationship between the synchronously acquired stator phase-currents ia, ib, and ic and the corresponding d , q and 0 axis currents in rotor reference frame is given as shown below:

$$\begin{bmatrix} i_0 \\ i_d \\ i_q \end{bmatrix} = \left(\sqrt{\frac{2}{3}}\right) \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \cos(\theta) & \cos\left(\theta - 2\pi/3\right) & \cos\left(\theta + 2\pi/3\right) \\ \sin(\theta) & \sin\left(\theta - 2\pi/3\right) & \sin\left(\theta + 2\pi/3\right) \end{bmatrix} \begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix} \quad (1)$$

Or in equation form, the direct and quadrature axis current can be obtained from three phase stator currents ia, ib, and ic by using the following equations:

$$i_d = (2/3) \times [i_a \cos(\theta) + i_b \cos(\theta - 2\pi/3) + i_c \cos(\theta + 2\pi/3)] \quad (2)$$

$$i_q = (2/3) \times [i_a \sin(\theta) + i_b \sin(\theta - 2\pi/3) + i_c \sin(\theta + 2\pi/3)] \quad (3)$$

Under ideal conditions, for healthy motor, the plot between 'id' and 'iq' ha s a circular shape, centred at origin. In case of faulty motor the plot changes in shape and thickness because of harmonic present in the air gap flux due to fault.

In this work the park's pattern are obtained for both the condition of motor that is when motor is with and without fault during its operation. Two types of absorbability is studied in the work:

1. Unbalanced supply
2. Fault in ball bearing of the motor

Under both normal and abnormal conditions, the stator current signals were acquired from the motor and the Park patterns (i.e. plot between id and iq) were obtained and given below.

Figures 2(a) shows the Park's pattern for the sinusoidal supply fed cage induction motor drive under no fault condition. It is clearly shown that the pattern is circular in shape, though they obtained shape is not a crisp circle due to actual operating condition.

Figures 2(b) shows the Park's pattern for the sinusoidal supply fed cage induction motor drive with small fault introduced in one of the balls of bearing. The deviation in circular shape is clearly visible.

Figures 2(c) shows the Park's pattern for the sinusoidal supply fed cage induction motor drive with small fault introduced in outer raceway of the bearing. The deviation in circular shape is clearly visible.

Figures 2(d) shows the Park's pattern for the sinusoidal supply fed cage induction motor drive with unbalance of 10%, introduced in one of the phase.

Figures 2(e) shows the Park's pattern for the sinusoidal supply fed cage induction motor drive with unbalance of 40%, in one of the phase under full load condition

Figures 2(f) shows the Park's pattern for the sinusoidal supply fed cage induction motor drive with unbalance of 10%, in two phases under full load condition.

Figures 2(g) shows the Park's pattern for the sinusoidal supply fed cage induction motor drive with unbalance of 20%, introduced in one of the phase.

Figures 2(h) shows the Park's pattern for the sinusoidal supply fed cage induction motor drive with unbalance of 30%, introduced in one of the phase under full load condition.

VIII. CONCLUSIONS

In this paper, a fault detection system for induction motor was proposed which uses park transformation of the stator currents. It is found that the patterns generated by park vectors are distinct under healthy condition, with unbalance in supply side and with damaged bearing. Hence any abnormality in motor can be easily observed by the shape of Park's patter. However authors feel that more data are to be gathered to develop the comprehensive database in order to classify the abnormality developing in motor at their initial stages so that corrective measures can be taken in cost effective manner.

ACKNOWLEDGEMENT

The authors would like to thank AICTE, New Delhi for providing financial support under RPS Research Promotion Scheme

REFERENCES

[1] G.K Singh and S.A.S Al Kazzaz, ``Induction machine drive condition monitoring and diagnostic research-a survey,” Electrical Power Systems Research.vol.64,no.2, pp,145-158, 2003.

[2] W.T. Thomson and M. Fenger,`` Current signature analysis to detect induction motor Faults,” IEEE Industry Applications Magazine,vol.7,no.4.pp. 26-34,2001.

[3] H.Nejjari, M. Benbouzid, Monitoring and diagnosis of induction motors electrical faults using a current park’s vector pattern learning, approach, IEEE Transaction on Industry Applications 36 (3) (2000) 730-735

[4] A.J.M Cardoso and E.S. Saraiva,`` The use of the Park’s Vector in the detection of electrical failures on three-phase induction motors.” In Proc, Int AMSE Conf. Modelling.

[5] A.Bellini,F.Filippetti,G.Franceschini,C.Tassonian dG.B.Kliman,``Quantitative Evaluation of induction motor broken bars by means of electrical signatureanalysis,” IEEETrans.onInd. Appl.,Vol.37,no.5,Sep./Oct.2001,pp.1248-1255.

[6] Benbouzid MEH, Bibliography on induction motors faults detection and diagnosis. IEEE Trans. Energy Convers 1999;14(4):1065-74

[7] Benouzza. N, Benyettou. A, Bendiabdellah A. "An advanced Park’s vectors approach for rotor cage diagnosis." In: Proceeding. Of first international symposium on control, communication and signal processing, 2004.p.461-4.

[8] Haji M, Toliyat HA, “Pattern recognisation- a technique for induction machines rotor broken bar detection". IEEE Trans Energy Convers 2001; 16(4):312-7.

[9] Silva JLH, Marques Cardoso AJ "Bearing failures diagnosis in three-phase induction motors by extended park’s vector approach." In : proceeding of the IEEE IECON’05, Raleigh, NC,USA, November 2005.p.2591-6

[10] Zarei J. Poshtan J."An advanced Park’s vectors approach for bearing fault detection. " In: Proceedings of the IEEE international conference on industrial technology (ICIT), India, 15-17 December 2006.p.1472-9.

[11] Zarei J. Poshtan J, “A new method based on Park’s vectors approach for bearing fault detection.”In: Proceeding of 20th international power system conference, Tehran (Iran), 16-18 November 2005[in Persian language].

[12] W. T. Thomson, M. Fenger, "Current signature analysis to detect induction motor faults, IEEE Industry Application Magazine, July/August 2001,pp.26-34.

[13] G. B. Kliman, R.A.Koegl, J.Stein, R.D.Endicott, M. W. Madden, "Noninvasive detection of broken rotor bars in operating induction motors", IEEE Transactions on Energy Conversion EC-3 (4) (1986) 873-879.

[14] Peter Vas, “Parameter estimation, condition monitoring, and diagnosis of electrical machines”, Clarendon Press Oxford., 1993.

[15] J. Robinson, C. D. Whelan, and N. K. Haggerty, “Trends in advanced motor protection and monitoring,” IEEE Transactions on Industry Applications, Vol. 40, No. 3, pp. 853-860, 2004

[16] Jafar Milimonfared, Homayoun Meshgin Kelk, Subhasis Nandi, Artin Der Minassians and Hamid A. Toliyat, “A novel approach for broken-rotor-bar detection in cage induction motors”, IEEE Transactions on Industry Applications, Vol. 35, No. 5, pp.1000-1006, 1999.

[17] Joksimovic, G. M., and Penman, J., “The detection of inter-turn short circuits in the stator winding of operating motors”, IEEE Transactions on Industrial Electronics Vol.47, No. 5, October, pp. 1078-1084, 2000.

[18] Jung J.H., Lee J.J., and Kwon B.H., “Online diagnosis of induction motor using MCSA”, IEEE Transactions on Industrial Electronics, Vol. 53, No. 6, pp. 1842-1852 Dec. 2008.

[19] D. E. Schump, “Reliability testing of electric motors,” IEEE Transactions on Industry Applications, Vol. 25, No. 3, pp. 386-390, May/June, 1989

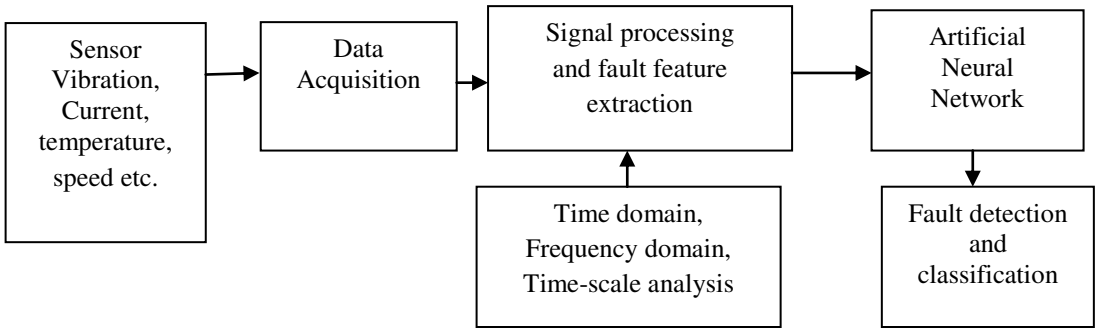


Figure 1 Signal processing approach to fault diagnosis

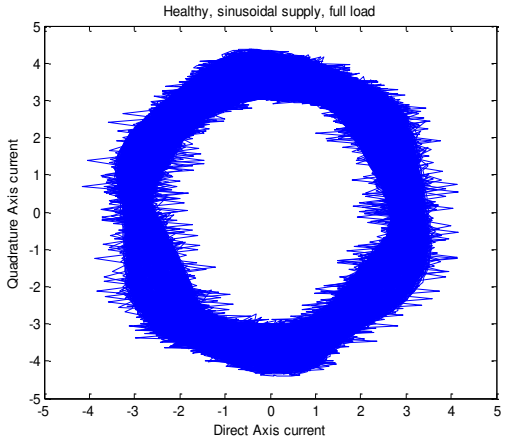


Figure 2(a): Park's pattern for healthy Induction motor at full load

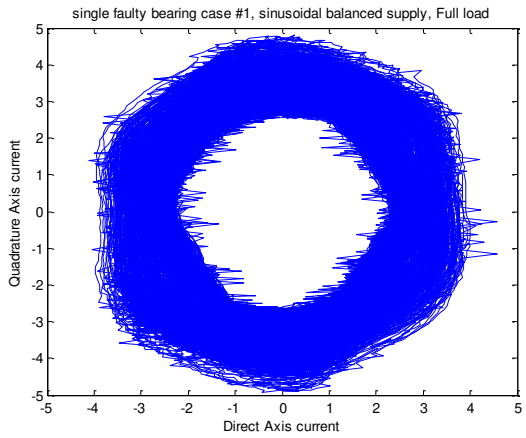


Figure 2(b): Park's vector for bearing fault of low severity under full load condition

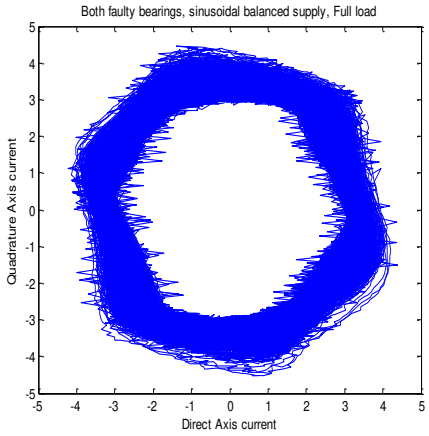


Figure 2(c): Park's vector for Bearing Fault for increased severity under full load condition

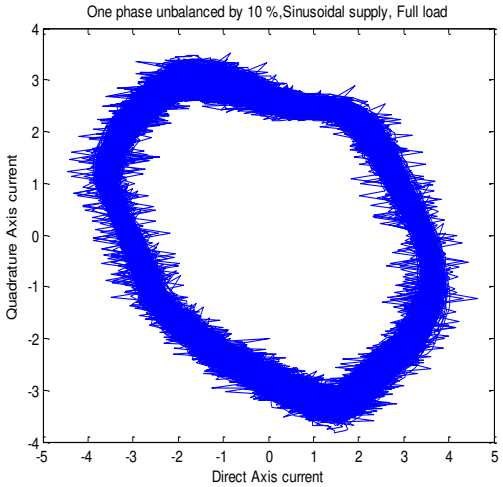


Figure 2(d): Park's vector for 10% unbalanced supply in one phase voltage under full load condition

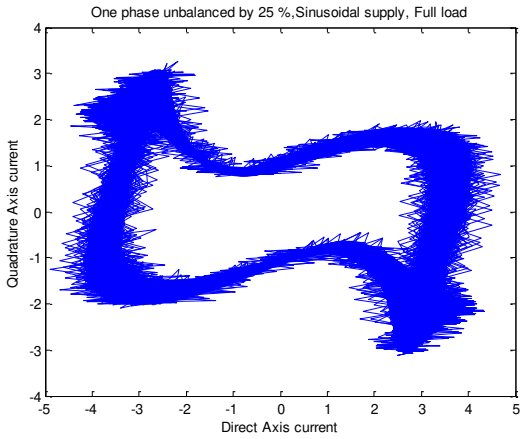


Figure 2(e): Park's vector for 40 % Unbalanced supply in one phase voltage under full load condition

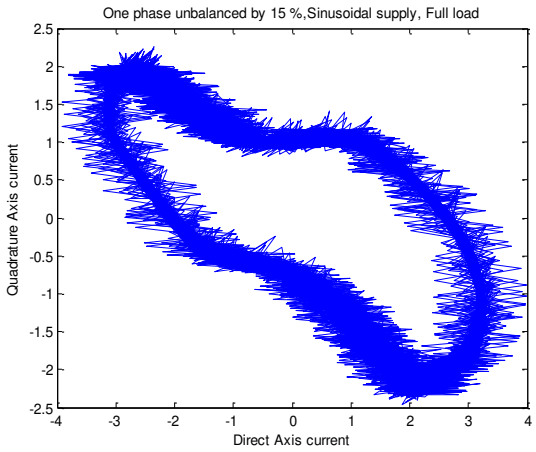


Figure 2(g): Park's vector for 20 % Unbalanced supply in one phase voltage under full load condition

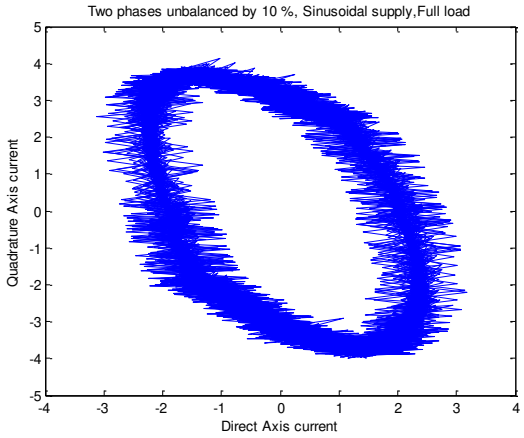


Figure 2 (f): Park's vector for 10 % Unbalanced supply in two phase voltage under full load condition

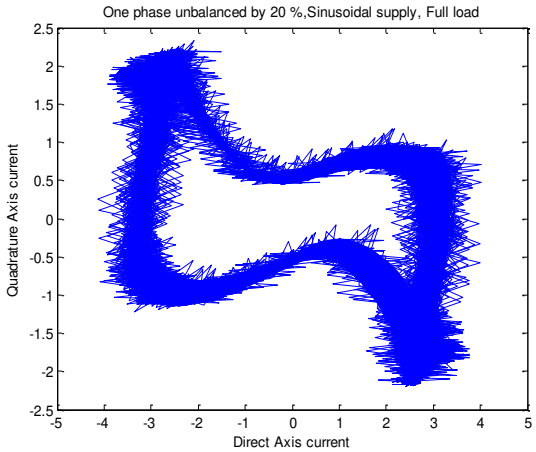


Figure 2(h): Park's vector for 30 % Unbalanced supply in one phase voltage under full load condition

Implimentation of A Parallel IEEE P754 Decimal Floating-Point Multiplier Using Verilog

Vrushali Ailawar

Dept of electronics Engg., Priyadarshini College of Engineering, Nagpur, India
E-mail : Vrushaliailawar@gmail.com

Abstract **-** Decimal multiplication is important in many commercial applications including financial analysis, banking, tax calculation, currency conversion, insurance, and accounting. This paper designs a fully parallel decimal floating-point multiplier that follows the recent draft of the IEEE P754 Standard for Floating-point Arithmetic (IEEE P754). The 754-1985 standard defines formats for representing floating-point numbers including negative zero and denormal numbers and special values infinities and Not-a-Number (NaN) together with a set of *floating-point operations* that operate on these values. The novelty of the design is that it is the first parallel decimal floating-point multiplier offering low latency and high throughput. Novel features of the multipliers include support for decimal floating-point numbers, on-the-fly generation of the sticky bit in the iterative design, early estimation of the shift amount, and efficient decimal rounding. The fixed-point design is extended to support floating-point multiplication by adding several components including exponent generation, rounding, shifting, and exception handling.

Keywords *—Decimal multiplication, Not A Number (NaN) ,binary coded decimal, floating-point arithmetic, serial multiplication, parallel multiplication*

I. INTRODUCTION

A fundamental operation in DFP arithmetic is multiplication, which is integral to the decimal-dominant applications found in financial analysis, banking, tax calculation, currency conversion, insurance, and accounting .Previous decimal multipliers have primarily focused on fixed point multiplication. Designs including use a sequential approach of iterating over the digits of the multiplier and selecting an appropriate multiple of the multiplicand. Generally, these designs have high latency and low throughput due to their sequential approach. This paper presents a parallel DFP multiplier based on a parallel fixed-point multiplier and a previous implementation of a DFP multiplier. The floating-point multiplier presented in this paper is based on the radix-10 fixed-point multiplier in due to its highly efficient structure. This multiplier generates a sufficient subset of multiplicand multiples and then selects all the partial-products in parallel based on the digits of the multiplier operand. To the best of our knowledge, this is the first published design of a parallel decimal floating-point multiplier that is compliant with IEEE P754.

II. BACKGROUND

Floating-point numbers are the favorites of software people, and the least favorite of hardware people.

The reason for this is because floating point takes up almost 3X the hardware of fixed-point math. The advantage of floating point is that precision is always maintained with a wide dynamic range, where fixed point numbers loose precision. The IEEE P754 standard specifies formats for both binary floating-point (BFP) and decimal floating-point (DFP) numbers .The primary difference between the two formats, besides the radix, is the normalization of the significands.BFP significant are normalized with the radix point to the right of the most significant bit (MSB), while DFP significands are not required to be normalized and are typically represented as integers. In this paper, all DFP operands use integer significands.The IEEE P754 standard specifies DFP formats of 32,64, and 128 bits. An IEEE P754 DFP number contains a sign bit, an integer significand with a precision of p digits, and a biased exponent. The value of a finite DFP number is:

$$D = -1s \times C \times 10E-bias \quad (1)$$

Where s is the sign bit, C is the non-negative integer significand,and E is the biased non-negative integer exponent[1]. IEEE P754 defines a preferred exponent, which for multiplication is:

PE = EA + EB − bias (2)

Where EA and EB are the biased exponents.

The multiplier uses the preferred exponent when encoding the result of a multiplication, so long as this does not lead to a loss of precision. Exponents of the first and second operands, respectively. DFP numbers with significands encoded in the DPD format. This format has p =1 decimal digits of precision in the significand, an unbiased exponent range of [−383, 384], and a bias of 398.

III. MULTIPLIER DESIGN

A general overview of our parallel floating-point multiplier design is presented in Figure 1. Arrows are used to indicate the direction of data flow and dashed lines separate the main stages of the design.

.

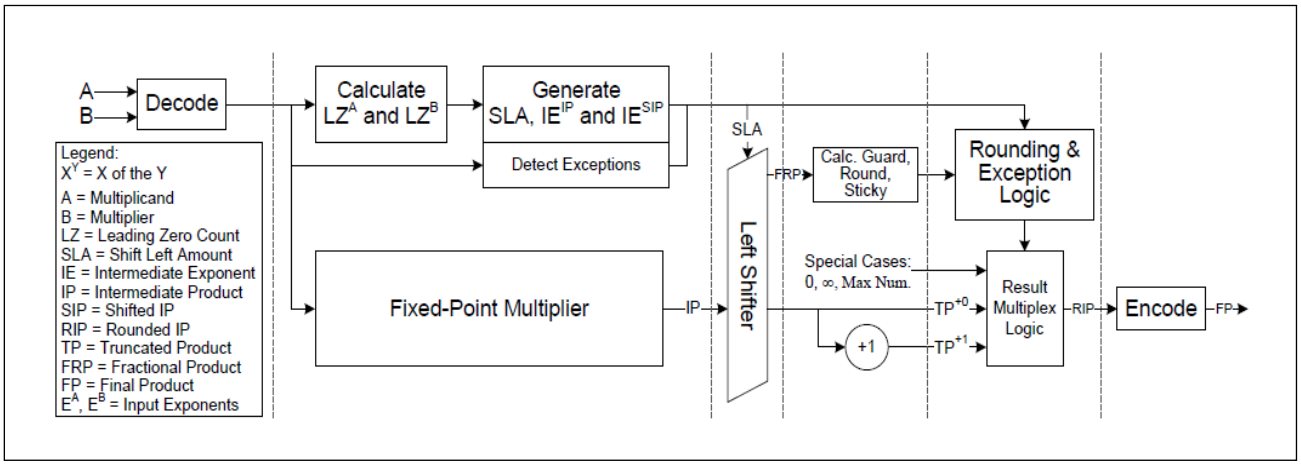


Figure: 1 High- Level Decimal Floating- Point Multiplier Diagram

The multiplication begins with reading two operands in IEEE P754 format and decoding each to produce the sign bit, significand, exponent, and flags for special values of Not-a-Number (NaN) or infinity. The significands of the two operands are then decoded from the DPD encoding to Binary Coded Decimal (BCD).

As soon as the decoded significands become available, a decimal fixed-point multiplication begins. If one or both of the operands is a NaN, its value is preserved through the multiplier by forcing the other operand to a value of one. The fixed-point multiplier generates 16 decimal partial products in parallel and adds them along with a possible correction term using a carry-save adder (CSA) tree followed by a high-speed decimal carry-propagate adder. The result is a non-redundant, 32-digit BCD number referred to as the intermediate product (IP).

In parallel with the multiplication, a shift-left amount (SLA) and corresponding intermediate exponent of the intermediate product (IE^{IP}) are calculated for shifting the 32-digit fixed-point result to fit into p = 16 digits of precision. This calculation is performed using leading-zero detection (LZD) on both operands in order to estimate the number of significant digits in the result.

Since the calculation occurs prior to the computation of the IP, the SLA may be off by one due to the significance of the product being one less than expected. In addition to the SLA and IEIP values, this unit calculates the sign bit of the final result and detects exception conditions which are later used by the rounding unit. In parallel with the multiplication, a shift-left amount (SLA) and corresponding intermediate exponent of the intermediate product (IE^{IP}) are calculated for shifting the 32-digit fixed-point result to fit into p = 16 digits of precision. This calculation is performed using leading-zero detection (LZD) on both operands in order to estimate the number of significant digits in the result. Since the calculation occurs prior to the computation of the IP, the SLA may be off by one due to the significance of the product being one less than expected. In addition to the SLA and IEIP values, this unit calculates the sign bit of the final result and

detects exception conditions which are later used by the rounding unit.

The 32-digit IP from the multiplier is then shifted by the SLA amount, forming the shifted intermediate product (SIP). The design sets the decimal point to be in the middle of the 32-digit intermediate product, thus splitting it into a 16-digit truncated product (TP^{+0}) and a 16-digit fractional product (FRP). This design choice keeps the decimal point in the same location throughout the data path and requires only a left-shift to produce the SIP.

Next, the FRP is used to produce the guard digit, round digit, and sticky bit. In parallel, the TP+0 is incremented to allow the rounding logic to select between TP+0 and TP+1, which is sufficient to support all rounding modes. Finally, the rounding and exception logic uses the rounding mode and exception conditions to select between TP+0, TP+1, and special case values to produce the rounded inter mediate product (RIP).

The RIP is then encoded in Densely Packed Decimal (DPD) and put in IEEE P754 format with the appropriate Flags set to produce the final product (FP). To illustrate the multiplication process, consider multiplying 3141592654 × 10−9 by 988121822 × 1011 When the rounding mode is round Ties To Even, as depicted in Figure 2. Our parallel multiplier starts by decoding the inputs from IEEE P754 format. Next, fixed-point multiplication is performed on the integer significands. In parallel, the SLA value of 13 is calculated by summing the leading zero counts of both operands and the unbiased version of the IE^{IP} value is calculated as 11 + (−9) + p = 18. After the fixed-point multiplication is complete, the result (IP) is shifted left by SLA = 13 digits to form the shifted intermediate product (SIP). The truncated integer (TP) portion of the result is then incremented, and rounding is performed to select between TP+0 and TP+1 as the final product (FP).

A. *Fixed Point Multiplier*

A key component of the design is the fixed-point multiplier, which is based on the radix-10 parallel fixed-point multiplier .The novelty of this fixed-point design is in using a special BCD digit recoding to reduce logic and in generating partial products in parallel. The fixed point multiplier unit takes two 16-digit integer operands, calculates 16 partial products in parallel and returns their sum, a 32-digit integer. There are three main components in the fixed-point multiplier design: generation of multiplicand multiples, selection of partial products, and reduction of partial products. The multiple generation stage produces {1x...5x} multiples of the multiplicand. Next, positive or negative versions of the multiplicand multiples are selected based on the Booth sign-encoded digits of the multiplier operand. All of the selected multiples (partial products) are then reduced using a carry-save adder (CSA) tree followed by a high-speed decimal carry-propagate adder. In the first stage of the process, {1x...5x} multiples of the multiplicand are generated. All of the resulting products are encoded in BCD-4221 to simplify the CSA tree.[ii] The 1x, 2x, 4x and 5x multiples are generated in a novel way using digit recoding and wired shifts, however, the 3x multiple requires a BCD adder to sum 1x and 2x.The next stage uses the digits of the multiplier to select, in parallel, 16 multiplicand multiples which form the partial products Signed-digit recoding is used to require only the{1x...5x} multiples and their complements. If a multiplier digit is greater than 5, a negative multiple is selected and 1x is added to the next digit (e.g. 7 = −3 + 10). In this way every multiplier digit selects ±{1x...5x}. As a special case, if the most significant digit (MSD) of the multiplier is greater than 5, then an extra corrective partial product of 1x is added to the partial product tree. After the partial products have been selected they are reduced using a CSA tree. Since all 16 combinations of the BCD-4221 encoded partial-products are valid decimal values, a slightly modified binary CSA tree is used .The resulting carry and Partial sum use the BCD-8421as shown in figure (3) encoding and are added using a high-speed direct decimal carry-propagate adder to produce a 32-digit BCD-8421 result. The direct decimal adder uses a Kogge-Stone network to quickly produce the carries between digits

B. *Intermediate Exponent and Shift Calculation*

Two key computations that occur in parallel with the Fixed-point multiplication is exponent calculation and shift left amount calculation. At the end of the fixed point multiplication, p digits of the intermediate product are to the right of the decimal point. This increases the intermediate exponent of the intermediate product (IE^{IP}) by p, giving the following equation for the biased intermediate exponent:

$$IE^{IP} = E^A + E^B - \text{bias} + p$$

$$= PE + p$$

Since there are significant digits to the right of the decimal point, the intermediate product produced by the fixed point multiplier may need to be shifted to achieve the preferred exponent or to bring the product exponent into range. The shift left amount is determined by first calculating the estimated number of significant digits in the intermediate product, S^{IP}, where S^{IP} is the sum of the number of significant digits of the two operands as given by

$$S^{IP} = S^A + S^B.$$

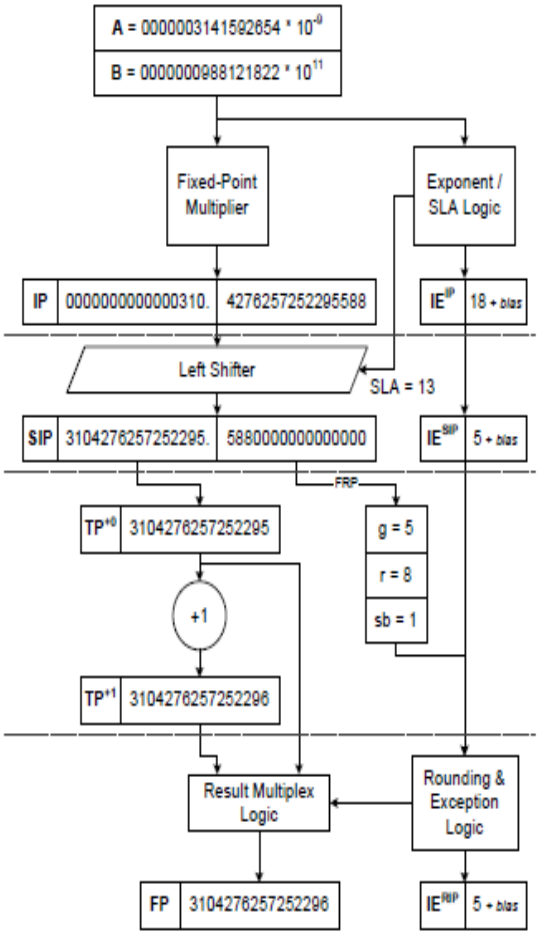


Figure: 2 DFT Multiplier Example

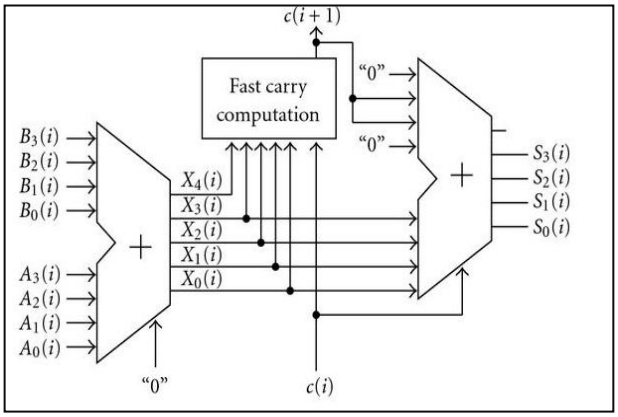


Figure 3: BCD-8421 full adder.

Next, two cases are considered. If $S^{IP} > p$ then some number of zeros need to be shifted off to the left in order maximize the significance of the product. If $S^{IP} \leq p$, then all of the significant digits of the product are to the right of the decimal point, and the shift value should be p, in order to place the result to the left of the decimal point. This leads to the following equations for the shift left amount or SLA.

$$SLA = \min((2p) - (S^A + S^B), p)$$

$$= \min((2p) - ((p - LZ^A) + (p - LZ^B)), p)$$

$$= \min(LZA + LZB, p)$$

After the left shift is performed, the intermediate exponent is decreased by the SLA. This intermediate exponent associated with the shifted intermediate product is calculated simply as

$$IE^{SIP} = IE^{IP} - SLA.$$

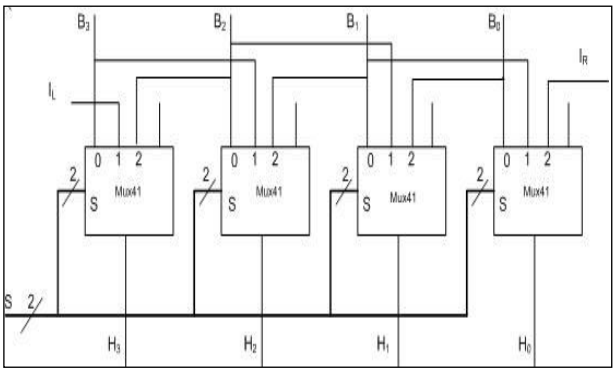


Figure 4 : 4bit shifter

C. *Overflow and Underflow*

Efficiently handling overflow and underflow in decimal multiplication provides a significant challenge and requires changes to SLA generation, exponent calculation, and rounding. For overflow, detection is done after rounding by comparing the intermediate exponent of the rounded intermediate product (IE^{RIP}) with the maximum exponent, E_{max}. During exponent and SLA generation,if $IEIP - SLA > E_{max}$, then our design attempts to avoid overflow by increasing SLA in order to decrease IESIP . The SLA value can only be increased to the point where all leading zeros of the intermediate product have been removed. That is, the following must hold:
$SLA_{new} \leq (LZA + LZB)$. If the maximum SLA yields $IE^{SIP} = E_{max} + 1$, then it is still possible to avoid overflow if a corrective left shift is performed during rounding. If after rounding $IE^{RIP} > Emax$, then overflow has occurred and an overflow value based on the sign and rounding mode is selected. as shown in Table 1. In parallel with fixed-point multiplier using the following equation.

$$SRA = \min((Emin - IE^{IP}), p + 2)$$

In this case, IE^{SIP} must also be modified to be IE^{IP} +SRA when a subnormal result is detected. The same conditions can still be used to detect underflow.

D. Rounding

Rounding must be performed when all of the significant digits of the intermediate product do not fit within the p. digits of the result's significand. The IEEE P754 draft standard specifies several rounding modes that must be supported. Each rounding mode and its associated condition(s) are listed in Table 1. In this table, the ^ and v symbols represent the logical OR and AND operations respectively.

In the normal case without overflow or underflow, rounding is accomplished by selecting either the shifted intermediate product (SIP) truncated to p digits or its incremented value. These values are represented as TP^{+0} and TP^{+1}, respectively. To perform the rounding selection, several flags and indicators must be calculated. These values are all calculated in parallel with the incrementer to improve the latency of the overall design. First, the sticky bit, sb, is calculated from the lower 14 digits of the SIP using a standard tree of OR gates. Next, an indicator, grsb, is set whenever there are non-zero digits to the right of the decimal point. This is calculated as:

= (g > 0) _ (r > 0) _ sb.

Finally, the guard digit must be incremented and the possible carry-out of the guard digit indicated to the rounding selection logic. The incremented guard digit is used in the case of both a corrective left shift and round-up. A one-digit direct decimal adder is used to perform this increment and the carry-out is used as the flag g9 to indicate that a carryout as occurred. Using the values for TP+0, TP+1, rucls==0 , rucls==1 , grsb, g, g+1, and g9, the algorithm presented in Figure 5 is used to realize IEEE P754 compliant rounding.

Table 1. Rounding Mode, Round Up Conditions, and Product Override for Overflow

Rounding Mode	Condition for Selecting TP^{+1} (Non-overflow)	Product Override (Overflow) $s^P == 0$	$s^P == 1$
roundTiesToEven	$(g > 5) \vee ((g == 5) \wedge (l \vee (r > 0) \vee sb))$	$+\infty$	$-\infty$
roundTiesToAway	$g \geq 5$	$+\infty$	$-\infty$
roundTowardPositive	$!s^P \wedge ((g > 0) \vee (r > 0) \vee sb)$	$+\infty$	−Max Num.
roundTowardNegative	$s^P \wedge ((g > 0) \vee (r > 0) \vee sb)$	+Max Num.	$-\infty$
roundTowardZero	*none*	+Max Num.	−Max Num.

Legend: *g*=guard digit, *r*=round digit, s^P=sign of the product, *sb*=sticky bit, *l*=LSB of TP^{+0}

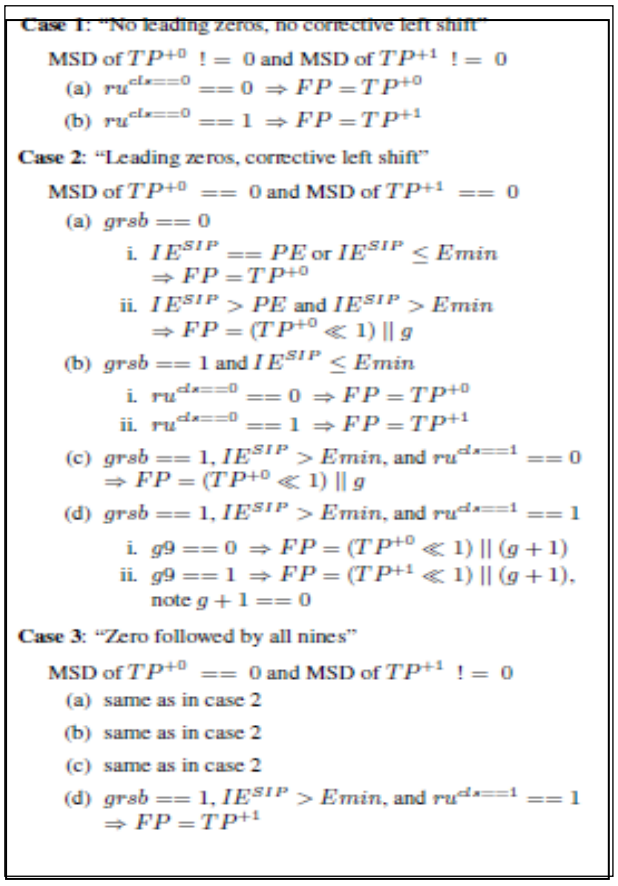


Figure 5: Rounding Scheme

IV. COMMON APPLICATION:

The advantage of floating-point numbers is that they can represent a much larger range of values. In a fixed-point number representation, the radix point is always at the same location. While the convention simplifies numeric operations and conserves memory, it places a limit on the magnitude and precision of the number representation. In situations that require a large range of numbers or high resolution, a relocatable radix point is desirable. Very large and very small numbers can be represented in a floating-point format. Multiplication of floating-point numbers is also commonly required in DSP-based applications.

REFERENCES

[1]. A Parallel IEEE754 Decimal Floating Point Multiplier Brian Hickmann, Andrew Krioukov, and Michael Schulte University of Wisconsin-Madison{bjhickmann, krioukov, and Schulte}@wisc.edu

[2] Decimal Floating-Point Multiplication Mark A. Erle, Senior Member, IEEE, Brian J. Hickmann,

Member, IEEE,and Michael J. Schulte, Senior Member, IEEE

[3] IEEE Standard 754 Floating Point Numbers Steve Hollasch / Last update 2005-Feb-24

4] G. Bohlender and T. Teufel. Computer Arithmetic: Scientific Computation and Programming Languages, chapter BAPSC: A Decimal Floating-Point Processor for Optimal Arithmetic, pages 31–58. B. G. Teubner, 1987.

[5] M. S. Cohen, T. E. Hull, and V. C. Hamacher. CADAC: A Controlled-Precision Decimal Arithmetic Unit. IEEE Transactions Computers, C-32(4):370–377, April 1983.

[6] M. A. Erle and M. J. Schulte. Decimal Multiplication via Carry-Save Addition. In International Conference on Application-Specific Systems,

[7] IEEE Standards Committee. IEEE Standard for Floating-Point Arithmetic. World Wide Web. http://754r.ucbtest.org/drafts/754r.pdf.

[8] A. Vazquez, E. Antelo, and P. Montuschi. A New Family of High–Performance Parallel Decimal Multipliers. In 18th IEEE Symposium on Computer Arithmetic, pages 195–204, June 2007.

◈◈◈

[i] Biased exponent in this paper represented by E relate to IEEE P754's exponents by: E = e − (p − 1) + bias where e is the unbiased exponent as defined in the IEEE P754 Draft Standard.

[ii] In this paper, we represent alternate decimal encodings in the format BCD-xxxx where the x's are the weights of each binary bit. For example, 10012 has a value of 4+0+0+1 = 5 with the BCD-4221 encoding and a value of 8 + 0 + 0 + 1 = 9 with the BCD-8421 encoding.

All Digital Phase Locked Loop (ADPLL) For High Performance Microprocessors

P. Bosebabu & K. Rakesh
MVGR College of engineering/Electronics and Communication Engineering, Vizianagaram, A.P., INDIA
E-mail: pbosebabu@gmail.com, rakesh_mvgr@yahoo.com

Abstract - The phase locked loop has become a common place in high performance microprocessors in which the tasks of frequency synthesis,dutycycle enhancement and clock deskewing are taken care.The analogPLL can also do the above tasks but it suffers from delay in acquiring phase and frequncy which can be overcomed by DPLL(digital phase locked loop).The ADPLL(all digital phase locked loop)has characterstic of very fast lock in time which is suited for clock generation on high performance microprocessors like PowerPC supporting fast entry and exit from power mangement techniques. A frequency-synthesizing, all-digital phase-locked loop (ADPLL) has a gain mechanism independent of process, voltage, and temperature, and is immune to input jitter. In the proposed model of ADPLL,the digitally-controlled oscillator (DCO) forms the core of the ADPLL and operates at the integer multiple of reference clock frequency .The DCO block is designed with continuouse time VCO and square wave converter.With the use simulink tool the ADPLL designed.

Keywords — DCO,continouse time VCO,Block processing block,SIMULINK tool.

I. INTRODUCTION

THE PLL has become commonplace in high-performance microprocessors, performing the tasks of frequency synthesis, duty-cycle enhancement, and clock de-skewing.The analog PLL traditionally fills this niche. Most all-digital phase-locked loops (ADPLL's) do not provide true frequencysynthesis; rather, they require a high-frequency clock source.This constraint mainly restricts their use to digital communications. However, the digital nature of the ADPLL makes it possible to achieve very fast lock times. This characteristic is attractive for clock generation on high-performance microprocessors, supporting fast entry and exit from powermanagement techniques.

This paper describes an innovative, frequency-synthesizing ADPLL that is fully integrated with a CMOS microprocessor. The performance metrics of the ADPLL rival or exceed those of prior analog implementations.The noteworthy performance metrics include: a deterministic 50-cycle phase lock, oscillator frequencies in excess of 900MHz, peak-to-peak jitter under 125 ps, and clock-to-reference skew under 250 ps. Key elements of the implementation are a digitally-controlled oscillator (DCO) with 8 b of binarilyweighted control, and a gain mechanism that is independent of process, voltage,and temperature fluctuations. The ADPLL features enhanced stability, increased input-jitter immunity, and low-voltage operation.

There are several variations of PLLs. Some terms that are used are analog phase-locked loop (APLL) also referred to as a linear phase-locked loop (LPLL), digital phase-locked loop (DPLL), all digital phase-locked loop (ADPLL), and software phase-locked loop (SPLL).

- Analog or Linear PLL (LPLL):

  The elements of LPLL are Phase detector is an analog multiplier, Loop filter is active or passive and uses a Voltage-controlled oscillator (VCO).

- Digital PLL (DPLL):

  The elements of DPLL are digital phase detector (such as XOR, edge-trigger JK, phase frequency detector), numerically controlled oscillator (NCO) and may have digital divider in the loop.

- All digital PLL (ADPLL) :

  The elements of ADPLL are Phase detector, filter and oscillator are digital. Both Continuous voltage controlled oscillator and square wave converter are used as a digitally controlled oscillator (DCO).

  - Software PLL (SPLL):

Functional blocks of SPLL are implemented by software rather than specialized hardware.

II. EXISTING MODEL:

A frequency-synthesizing, all-digital phase-locked loop (ADPLL) is fully integrated with a 0.5 pm CMOS microprocessor. The ADPLL has a 50-cycle phase lock, has a gain mechanism independent of process, voltage, and temperature, and is immune to input jitter. A digitally-controlled oscillator (DCO) forms the core of the ADPLL and operates from 50 to 550 MHz, running at 4x the reference clock frequency. The DCO has 16 bit binary weighted control and achieves LSB resolution under500 fs.

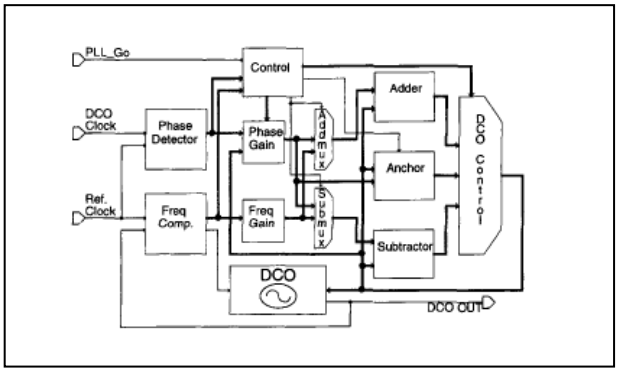


Figure 1: Existing model of ADPLL**.**

This is the model which is implemented in 1995, this is having the range from 50to 550Mhz,to advance this range , the proposed model is used to increase the range of frequency and also this model is used as a generalized model of ADPLL, with use of simulink tool.

III. PROPOSED MODEL OF ADPLL:

The block diagram of proposed model is shown below:

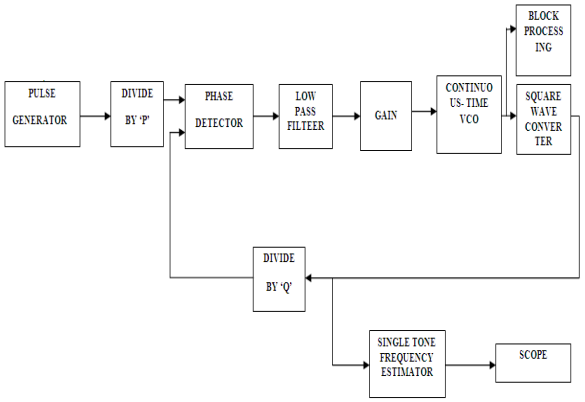


Figure 2: Block diagram of ADPLL

An all digital Phase Locked Loop is a closed loop control system which is used for the purpose of synchronization of the frequency and phase of a locally generated signal with that of an incoming signal. There are basically three components in a PLL. The components f ADPLL are the Phase detector (PD), the loop filter and the digitally Controlled Oscillator (DCO). The DCO is the heart of an ADPLL. An exclusive OR gate for serving the purpose of the PD and a first order filter for the loop filter and a typical DCO. Though the name Digital is present in the DPLL, it's not exactly a complete Digital PLL. The All Digital PLL makes an attempt at digitizing all the three components required for the operation of a phase locked loop. In its general form ADPLL consists mainly of a frequency generator (DCO), phase-frequency comparator, and a filter interconnect as shown in the figure 2 .

A reference clock f_{ref} is divided by an integer q, producing a signal $\frac{f_{ref}}{q}$ to which the DCO is to be locked. The function of the DCO is to generate a range of frequencies from a system clock which operate at a frequency f_s, much higher than the DCO output frequency f_{DCO}. Before being fed back to the phase comparator, the DCO output is divided by an integer p, so that the frequency $\frac{f_{DCO}}{p}$ is actually compared with $\frac{f_{ref}}{q}$. The result of this comparison is low pass filtered and fed back to the DCO, which adjusts the output frequency to reduce the error. The low-pass filter is a critical element in determining important parameters as acquisition time, jitter, and stability. Assuming proper design, the feedback mechanism converges and the DCO output frequency reaches steady state.

$$\frac{f_{DCO}}{p} = \frac{f_{ref}}{q} \rightarrow f_{DCO} = \frac{p}{q} f_{ref}.$$

IV. DIGITALLYCONTROLLED OSCILLATOR (DCO):

A digitally-controlled oscillator (DCO) is a heart of the ADPLL. The DCO consists of a frequency-control mechanism within an oscillator block. In the DCO design, the components used are continuous time voltage controlled oscillator (vco), converter and single tone frequency estimator .The vco produces signals based on the control signal, those signals are converted into square wave, and through the use of single tone frequency estimator, estimating the frequency of the signal.

Continuous time voltage controlled oscillator generates a signal with a frequency shift from the quiescent frequency parameter that is proportional to the input signal is interpreted as a voltage .if the input signal is U(t), then the output signal is

$$Y(t)=A_C \cos (2*\Pi*f_C*t+2*\Pi*k_C\int_{(0\ T)}U(T)dt_+\Phi) \quad (7)$$

Where A_C is output amplitude, f_C is quiescent frequency parameter, k_C is input sensitivity parameter. The continuous time voltage controlled oscillator uses continuous time integrator. The input and output both are sample based scalar signals. The quiescent frequency is defined as the voltage controlled oscillator output signal frequency without input.

The square wave converter is used for to convert the sampled signal produce by the continuous time VCO into square wave.

The single tone frequency estimator is used to estimate frequency of DCO.

The result of simulink model of DCO is shown in figure 9.

V. PHASE DETECTOR:

Left Phase detectors for phase-locked loop circuits may be classified in two types Type I detector is designed to be driven by analog signals or square-wave digital signals and produces an output pulse at the difference frequency. The Type I detector always produces an output waveform, which must be filtered to control the phase-locked loop voltage-controlled oscillator (VCO). A type II detector is sensitive only to the relative timing of the edges of the input and reference pulses, and produces a constant output proportional to phase difference when both signals are at the same frequency. This output will tend not to produce ripple in the control voltage of the VCO. The digital phase detector is produces less noise compared to analog phase detector, hence digital phase detectors like Ex-or gate is used.

The simulink model of phase detector output is shown in figure6.

VI. BUTTERWORTH FILTER:

The low pass filter is used for removing the high frequency and dc components of signal produced by the phase detector in phase locked loop. It is a critical element in the design of phase locked loop because, it determines the parameters jitter and stability of phase locked loop. In the design of ADPLL, the first order butter worth low pass filter in steady state analysis, then it can be act as a digital filter. It is having the advantage of maximally flat magnitude response. it is implemented in digital domain (i.e. in state space model).

The simulink model of Butterworth filter output is shown in figure7.

VII. FREQUENCY DIVIDER:

The frequency divider circuit produces the frequency divided signal of the applied signal. It can be designed through the D-flip flop. The D-type logic flip flop is a very versatile circuit. It can be used in many areas where an edge triggered circuit is needed. In one application this logic or digital circuit provides a very easy method of dividing an incoming pulse train by a factor of two.

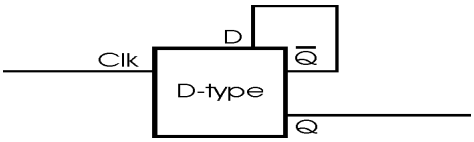


Figure 3: D-type frequency divide by two circuit.

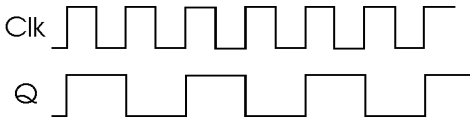


Figure 4: D-type frequency divide by two circuit wave forms.

The block processing block is used for checking the stability of ADPLL. It consists CMOS AND gate ,it takes one input from continuous time VCO in normalized matrix form , the another input is logic ,if it produces two synthesized signals as ‘o’ .hence the ADPLL is producing the stabled signal in terms of frequency and phase. These are called frequency maintenance and phase maintenance modes. Whenever CMOS AND gate produces the zero signal at the output side then two modes are operating correctly.

The simulink model of frequency divider block output is shown in figure5 and 8, figure 5 represents the frequency division of reference clock, and figure 8 represents frequency division of DCO output.

VIII. BLOCK PROCESSING BLOCK:

It is a free defined block in the simulink tool. Through the use of block processing block, the stability of the ADPLL is checking. It is taking the input from the continuous time VCO in normalized matrix form, when ever the all the elements in the normalized matrix are zero. Then the outputs of the block processing block is zero, it indicates that the ADPLL is stable in both frequency and phase .to see the phase maintenance and phase maintenance modes separately , then take two outputs from the block processing block. The internal

circuit of block processing block is 2-input AND gate , one input is normalized matrix and other input is not connected then when ever, the one input AND gate zero, then the output of AND gate is zero, without considering the second input.

The simulink model of block processing block outputs are shown in figures 10 and 11, figure 10 represents the frequency maintenance, and figure 11 represents phase maintenance of ADPLL.

IX. RESULTS:

The results for reference clock frequency 900MHZ with multiplication factor 0.25 is shown below:

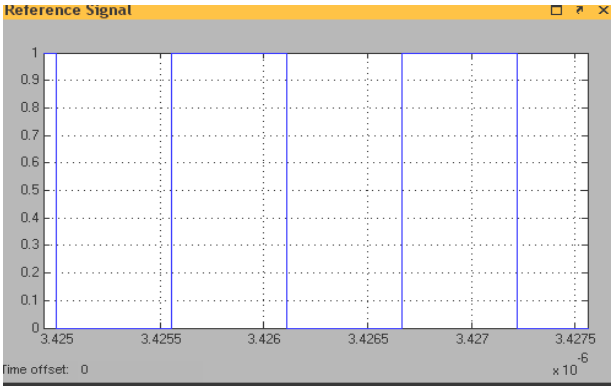


Figure 4: Reference clock pulse.

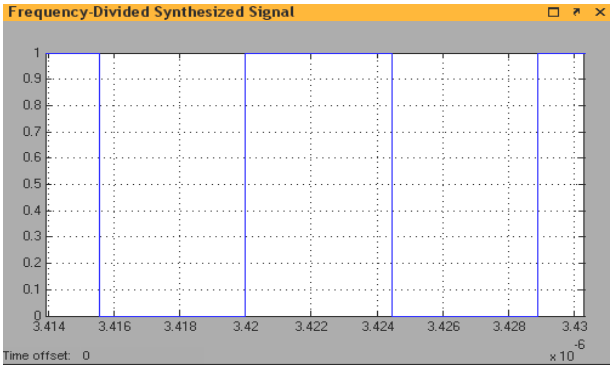


Figure 5 : Frequency division by 2.

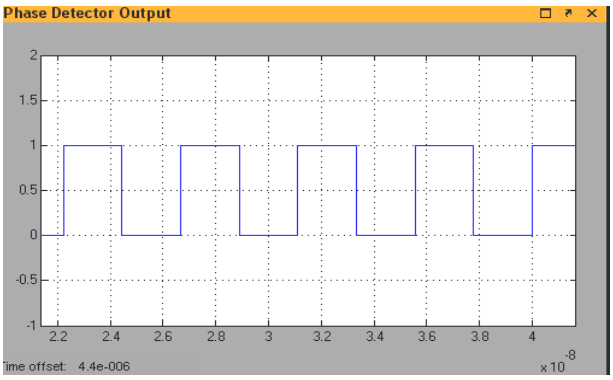


Figure 6: Phase detector output

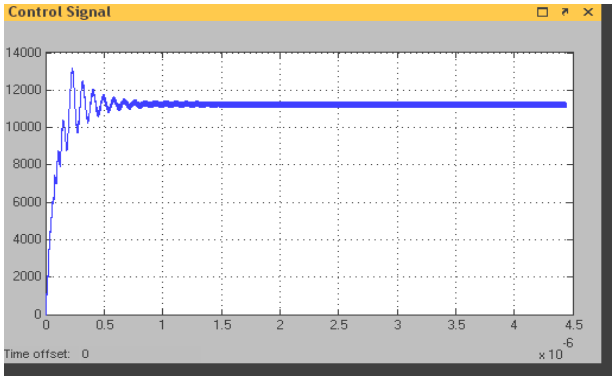


Figure 7: Control signal

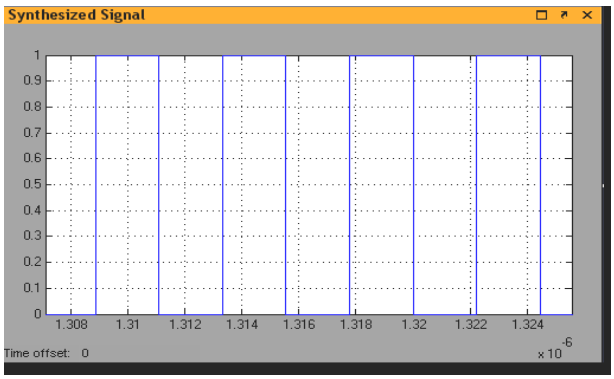


Figure 8: Square wave output

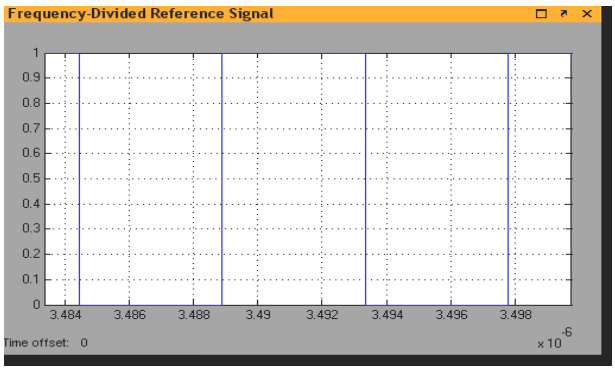


Figure 9: Frequency division by 8.

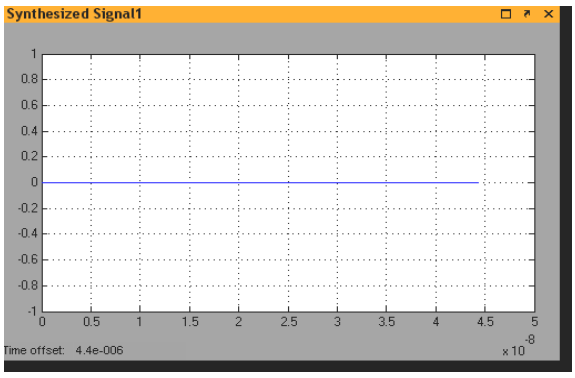


Figure 10: Synthesized signal1

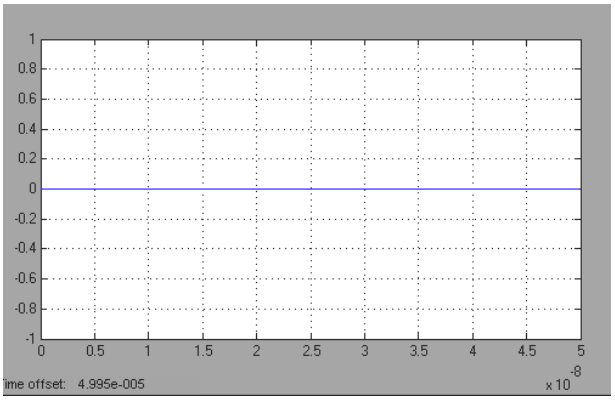


Figure 11: Synthesized signal2

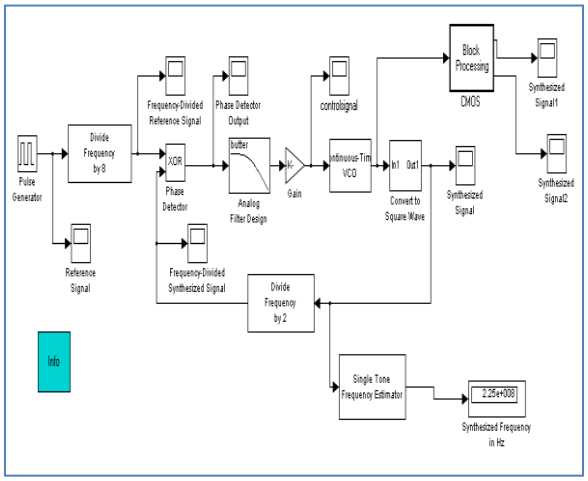


Figure 12: The synthesized frequency.

X. CONCLUSION:

A unique ADPLL implementation matches or exceeds the performance of traditional analog phase-locked loops. The DCO is the heart of the ADPLL, achieving performance and resolution comparable to state-of-the-art voltage-controlled oscillators. Because the DCO range (50-900 MHz) is a design constraint rather than a circuit constraint, the ADPLL has significant headroom. This ADPLL currently operates with this level of performance in a high-volume, high-performance, and low-power microprocessor. The proposed model of ADPLL is also be used as generalized model of ADPLL, because it is implemented at block diagram level.

ACKNOWLEDGMENT

We wish to thank to head of the department, technical and non-technical staff of electronics and communication engineering, and **MANASSAS TRUST** for their valuable suggestions and constant motivation that greatly helped the project to get successfully completed.

REFERECES

[1]. A PLL clock generator with 5 to 110 MHz lock range for microprocessors by Young er al., in ZSSCCDig. Tech. Papers, Feb. 1992.

[2]. Phase-Locked Loops, Theory, Design and Applications. by R. E. Best, New York: McGraw-Hill, 1993, 2nd ed.

[3]. A thesis on STUDY OF VOLTAGE-CONTROLLED OSCILLATOR BASED ANALOG-TO-DIGITAL CONVERTER by prof. k.r. subhashini, 2011.

[4]. MODERN VLSI DESIGN IP BASED DESIGN by Wayne wolf New York: Prentice Hall Modern Semiconductor Design Series,2009, 4th ed.

[5]. Simulink Basics Tutorial.

◈◈◈

Comparative Analysis of Image Compression Using Image Interpolation And RICE Codes With Quadtree Encoding

Vaibhav Jain, Nitin Jain, Ashish Parikh

Department of Electronics and communication Engineering, MIT Mandsaur
E-mail : vaibhav.jain@mitmandsaur.info, nitin.jain@mitmandsaur.info, ashish.parikh@mimandsair.info

Abstract - In this paper we have discussed comparative analysis of Image Compression Using Image Interpolation and RICE Codes with quad tree encoding. Firstly the original image is converted in to low resolution (less size) image, then quantization and quad tree encoding, techniques are applied to compress the image. To restore the original image, the compressed image is decompressed (dequantized & decoded) and interpolation techniques (tukey window and PG algorithm) are applied to achieve high resolution decompressed image. The PSNR of the high resolution decompressed image is high compared to PSNR of low resolution original image. At last we found the compression ratio is directly proportional to PSNR.

Keywords—SPIHT, ROI Compression, MIC, Peak Signal to Noise Ratio (PSNR), Compression Ratio (CR), Tukey-wind, PG algorithm (Papoulis-Gerchberg method).

I. INTRODUCTION

Image compression is the process of identifying and removing redundant data from images and then encodes it with a lossless or lossy encoder in such a way that the information is represented with less number of bits than the original image. Uncompressed high resolution images take lot of memory. For example, a single small 4″ × 4″ size colour picture, scanned at 300 dots per inch (dpi) with 24 bits/pixel of true colour, will produce a file containing more than 4 megabytes of data. This picture requires more than one minute for transmission by a typical transmission line (64k bit/second ISDN). Therefore large image files remain a major bottleneck in a distributed environment. This problem is generally overcome by various image compression techniques like JPEG compression. From the discussion it is clear that an image compression is essentially a quantization technique. Therefore at the receiver when the image is represented, there will be information loss which affects the overall resolution and clarity of the images. This is a major problem in medical images. With increasing number of medical image acquisition every day, storage, transmission and information preservation in such images has become a huge challenge. In this work we addressed this issue by proposing a novel technique for medical image compression and by showing the efficiency of the same.

Medical images are very sensitive to losses as the details are critical in terms of diagnosis and medical observation. Hence conventional image compression techniques are avoided in medical images. Medical Images need compression in such a way that medically relevant information (called ROI or region of interest) suffers minimum loss.

A variable bit stream coding is adopted for ROI image compression. A ROI coding is simply an image compression technique where different parts of the images are encoded at different rates, therefore having different compression ratio and PSNR. In a medical image, the area of the image over which the information is present is called ROI. In medical image compression scheme, a ROI can be determined by manual outlining or by segmentation technique. So the approach is to divide an image into regions and then apply encoding with different rates over transformed data of those images. So while encoding ROI part, quantization will consider more distinct levels where as non ROI part should be encoded by considering lesser number of quantization level.

The basis of any image compression is quantizing the values of the images and encoding the quantized values with the help of an encoding technique like entropy coding, Huffman coding, zero tree coding, SPIHT and so on. Quantization requires image values to be unique and need to remove the sparse information. Therefore representing the image in the transformed domain is important as against spatial domain as the significance of the pixels in spatial domain are not

known and therefore assigning them bits is difficult. Most widely used transformation for the image is wavelet transform.

A better choice is to first scale down the image resolution to an acceptable level (say 256x256), Apply compression and store or transmits. At the decoder, the image must be uncompressed and then should be mapped to a high resolution plane (1024x1024) using interpolation technique using Papoulis-Gerchberg method.

II. RELATED WORK

Ansari et al.[1] discusses about various medical image compression techniques and analyzes the performances. This work gives a significant insight into the region of interest compression.

For the compression of the images, it is essential to extract the uncorrelated information from the images and then encode the information using a suitable encoding technique. Zero Tree Encoding is one of the most used techniques in this direction and a low complexity variant is discussed by Debin [2] Ahmed and Ravi[3] introduces variable bit length coding in the Context of region of interest compression. They introduced a low rate encoding for region of interest and a high rate encoding for non region of interest.

Marco et al.[3] propose an image segmentation followed by rate adaptive wavelet based image coding. In this scheme the segmented areas are transformed using shape adaptive wavelets and are followed by separate encoding of each of the segmented parts separately. A PSNR value of average of 20-35db is achievable for .2bits par pixel encoding.

Lina [5] introduces a post processing technique for recovering the compression losses through local learning algorithm. This concept is quite helpful in ROI coding. Expecting certain compression errors in variable bit length coding; the degraded PSNR can be recovered using the technique proposed by [5].

Gibson [7] performs ROI compression in Angiogram video by first finding out the region of interest through obtaining correlated areas of frames and further applying wavelet based compression in ROI section. The work helps in explaining the concept of ROI in correspondence to medical images.

Christophe et al.[7] introduces 3-d SPIHT and also explains how variable bit rate can be allocated to different parts of images. The method also explains the mechanism of keeping the cut off point information through a global optimization.[8] Performs the ROI compression by dividing the image into two parts: ROI section and non ROI section. The ROI section is compressed with lossless integer wavelet and non ROI region is compressed with lossy wavelets to gain higher compression ration and also to maintain the PSNR.

[9] Proposes to extract the ROI region from the images based on gray scale texture coefficients and by clustering. Further the ROI regions are marked and both wavelet and Counterlet transform are used to generate the coefficients at the compression stage. The coefficients are encoded using both Huffman coding and modified EZW.

Marcus [10] presents a technique for ROI image compression using multiple decoders and presents visual results illustrating the final form of decoded image when the images are encoded with various bit planes.

The images here are encoded to multiple description sub images such that the final image can be reconstructed using one or more subset images.

In[11] Performs a variable bit rate encoding with forward encoding technique. In this method, the images are transformed and the bits of a section are encoded till a best rate is achieved. The remaining bits of the plane are left uncompressed. The remaining part is further transformed and the process is continued till it covers the entire image.

Cuhadar In [12] introduces a new concept for ROI coding called unbalanced ZeroTree. The observation finds that if a region of relatively small size is encoded, the region's wavelet coefficients are localized over a very small portion of the Zero-Tree. In this case, if we enable the use of nodes from lower levels of Zero-Trees, i.e., nodes other than roots of Zero-Trees, upon LIS initialization, we can save the redundant bits transmitted to reach the sub-tree comprising the wavelet coefficients of the region from the root of the Zero-Tree. 2) If we can establish homogeneity in tree structures, i.e., if all nodes of the tree have a common label, then we do not need bookkeeping of sets descendent labels during encoding. Based on these observations

In [13] Achieves ROI compression by iteratively sub samples the images into temporally low pass and high pass frames. The technique is summarized as: Subsample low-pass frames at even frames to obtain temporal low-pass frames and subsample high pass frames at odd frames to obtain temporal high-pass frames. Further Perform spatial 2D SA-DWT transforms according to their spatial shapes for every temporally transformed frame

In [14] combination of the object-based DCT coding and the high performance of the set partitioning in hierarchical tree (SPIHT) coding is used. The modification of the sub-band image data in the wavelet domain is done based on the DCT transformation and the classification of the wavelet coefficients in the LL

sub-band. The modification process provides a new sub-band image data containing almost the same information of the original one but having smaller values of the wavelet coefficient.

In [15] Divides the image into two parts called significant and insignificant image based on their visual properties. The regions are encoded after wavelet transform with rate 60% and 80% subsequently for the significant and the insignificant region.

In [16] explains the Ricean Code based image compression techniue. [17][18][19] [20][21][22][23] [24] Describes various methodologies for high resolution imaging. The papers discuss various projection mechanisms for mapping low resolution plane to high resolution plane.

III. PROPOSED WORK

All paragraphs must be indented. All paragraphs must be justified, i.e. both left-justified and right-justified.

Scale down or resize the image

- ✓ Apply SPIHT or Quad tree encoding to the sub images or sub band images.
- ✓ Store number of levels of decomposition in the encoded data.
- ✓ Decode data using SPIHT decoding
- ✓ Apply Inverse transform

Use Projection technique to scale up the low resolution image to high resolution one.

IV. METHODOLOGY

A. *INFORMATION PRESERVATION*

In most of the literature it is claimed that the information retained in the ROI of the medical images must be very high. Generally BPP or bits per pixel is decided first and then a compression is performed over the original image. There is no suitable measure of the quality of image other than PSNR. PSNR is a global phenomenon and do not reveal the information preservation or the information homogeneity. How much information is relevant to the image is not decided in the compression stage. Therefore we propose this unique scheme for medical image compression which is based on the principal of determining what amount of information compromise is acceptable by analyzing PSNR at every level of decomposition and continuing the process till reconstructed image's PSNR is acceptable (which is considered as 30Db[25]).

Once the PSNR go below the acceptable level, the iteration or down sampling of the image is stopped. This sub sampled image with wavelet coefficients is selected for the encoding. SPIHT [26][27] based encoding is adopted for generating the bit stream from the decomposed image.

As the analysis stage is significant and takes a lot of processing cycle, lifting wavelets are used for the transformation. Lifting wavelets increases the speed of the processing to a great deal and compensates for the time consumed for check and verify in the recursion stage for image sub sampling [28].

B. *IMAGE PROJECTION AND INTERPOLATION*

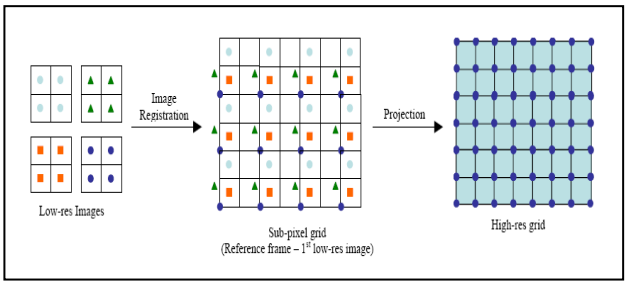


Figure 2: A Simple representation of image projection from low resolution image to high resolution image

Figure 2 describes the two steps graphically. On the left hand side four low-res images are shown (pixel values at integer pixel only). Motion estimation is used to estimate the pixel positions of the three images with respect to the 1^{st} image. Note that these pixel positions would not be at integer values, but can take any real value. Once this information is calculated accurately, it is possible to project this information on a desired high-resolution grid.

Before delving into the solution, lets first try to understand what are we trying to achieve. Let's say we start with a high-resolution image. Then we down sample it (without low-pass filtering). So, we have aliasing in the down sampled images. Now, in real-life we will be given these down-sampled images with aliasing and our goal is to create an HR image from them. So, basically as shown in the figure below, it boils down to finding the lost high-freq components and fixing the aliased low-freq values in the HR image. The Papoulis-Gerchberg algorithm (discussed in more detail below) along with POCS is used to solve this problem. It concentrates mainly on fixing the low-freq components, but it turns out that this algorithm also gives us some high-freq components depending on the information in the LR images.

The two steps used to solve this problem are shown below. The details of each step is given in the following sections

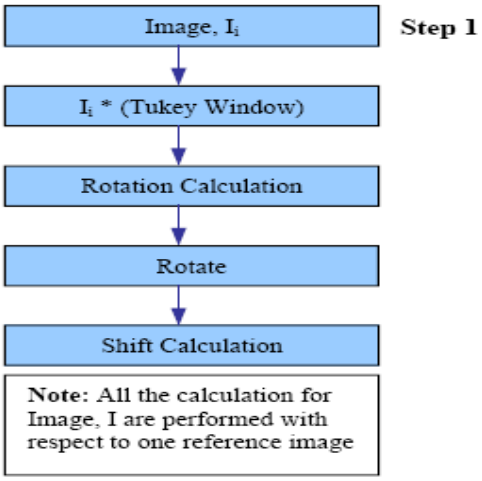


Fig. 3.Various operations done on the image before applying P-G Algorithm in step 1

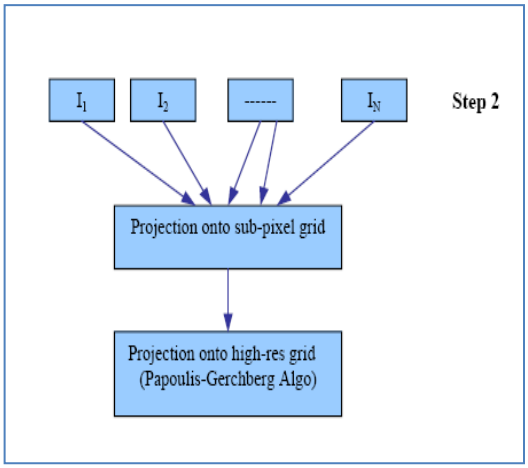


Fig. 4. Various operations done on the image after step 1

First an estimate of the rotation (mosaic) effect to be given to the given pixels for obtaining the super-resolution is obtained for the lower resolution data.

Rotation Calculation

There are two ways that can be used to carry out rotation estimation between low-res images.

- Rotate individual images at all the angles and correlate them with the first image. The angle that gives the maximum correlation is the angle of rotation between them.
- The method described above is computationally expensive, so a faster method is to calculate the image energy at each angle. This will give a vector containing average energy at each angle. The correlation of these energy vectors of different images will give the angle of rotation. The energy calculation could also be carried in frequency domain. A graphical representation below helps to illustrate the method clearly.

Angle (i) = max index (correlation ($I_1(\theta)$, Ii (θ)))

……. (17)

Where $I_1(\theta)$ is the pixel intensity of the reference pixel and Ii(θ) is the intensity of the i^{th} pixel. It turned out that though the first method is computationally expensive, but gives more precise results, so it was used in this work.

Shift Calculation

Once rotation angle is known between different images, shift calculation can be performed. Before calculating the shift, all the images are rotated with respect to the first image. Block matching was not used here for two reasons – 1. It is computationally expensive. 2. We need sub-pixel accuracy that is very hard to get with block matching. A much faster and reliable way to achieve this is using frequency domain method.

For determining the amount of shift in any pixel of an image, we use the relation

$F_i(u^T) = e^{j2\pi u \Delta s}$.

$F_1(u^T)$…… eq. (18)

This is obtained by applying Fourier Transform of a reference pixel matrix.The shift angle Δs from the above relation can be calculated as

$\Delta s = [\text{angle}\ (F_i(u^T)/ F_1(u^T))]/2\pi$ eq. (19)

And in matrix form, $\Delta s = [\Delta x\ \Delta y]^T$

…… eq. (20)

WhereU(x, y) is the pixel coordinate

Δx is the variation of current x-position from reference x- position

Δy is the variation of current y-position from reference y- position

$F_i(u^T)$ is the transform of transposed i^{th} pixel

$F_1(u^T)$ is the transform of transposed reference pixel

Δs is the shift angle

Technically, each frequency bin should give the same shift, but that does not happen. So, a least square method is used to estimate the shift Δs. As the LR images could

be aliased, only 6% of the lower frequencies were used to calculate Δs.

Projection

Once we have a good estimate of shift and rotation between all the images and the reference image, we can find the pixel positions of all the LR images in the reference frame of the first image. Then we can project this information on high-res grid. There are various techniques proposed for this. In this work we use Papoulis-Gerchberg algorithm[29][30][31] which is summarized as bellow.

This method assumes two things:

- Some of the pixel values in the high-res grid are known.
- The high frequency components in the high-res image are zero.

It works by projecting HR grid data on the two sets described above.The steps are:

- Form a high res grid. Set the known pixels values from the low-res images (after converting their pixel position to the ref frame of first low-res image). The position on the HR grid is calculated by rounding the magnified pixel positions to nearest integer locations.
- Set the high-freq components to zero in the freq domain.
- Force the known pixel values in spatial domain. Iterate.

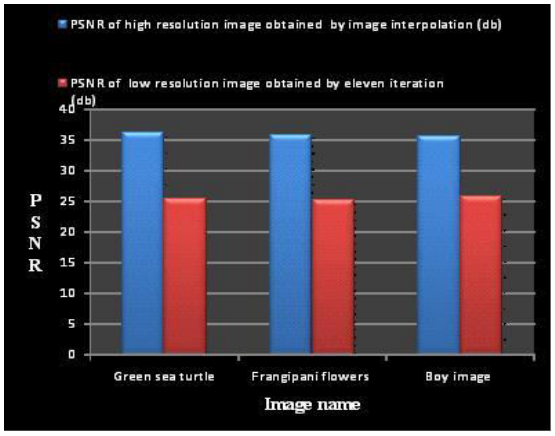
  By making the high-freq equal to zero, this method tries to interpolate the unknown values and so correct the aliasing for low-freq components. Also, by forcing the known values, it does predict some of the high-freq values.

V. RESULTS & OBSERVATION

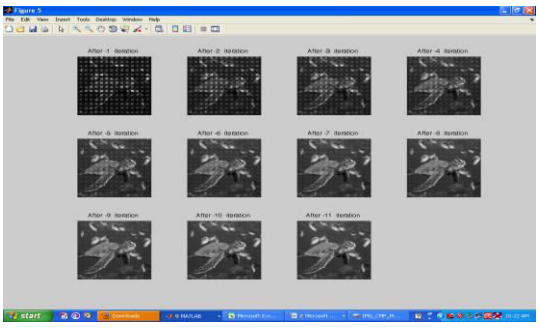
Experiments are perfomed on various images capture from mobile phones and digital camera and compare psnr of lowresolution image and high resolution image. In my technique first I convert original image in to low resolution then quantization and qadretree encoding will be done after then decoding of image will be done at receiver side I done eleven iteration for decompression of image to original image but at his stage image is in low resolution and has low psnr so I done something new to increase the psnr value at this stage so I use tukey window and pg algorithm to increase psnr value.

To prove my concept I perform operation on three image as shown in table

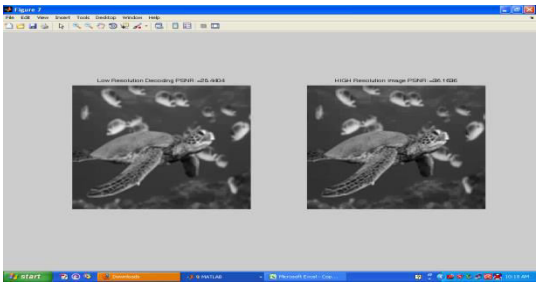
Name of the image	PSNR of high resolution decompressed image obtained by image interpolation (db)	PSNR of low resolution decompressed image obtained by eleven iteration (db)
Green sea turtle	36.1636	25.4404
Frangipani flowers	35.8863	25.1722
Boy image	35.6154	25.9674



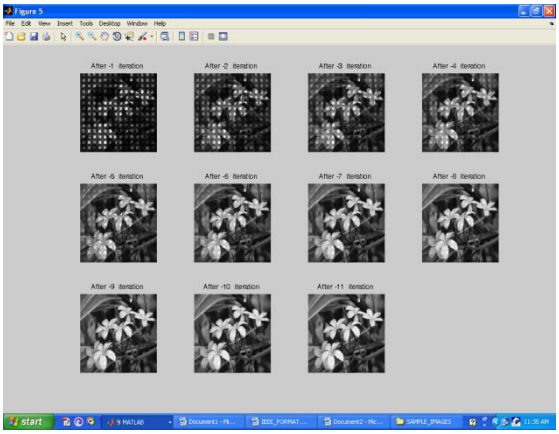
Graph construct in accordance with result analysis table



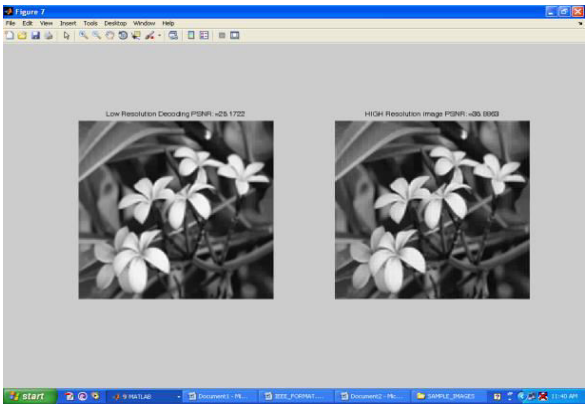
Green sea turtle Decompressed Low Resolution image after 11 iteration



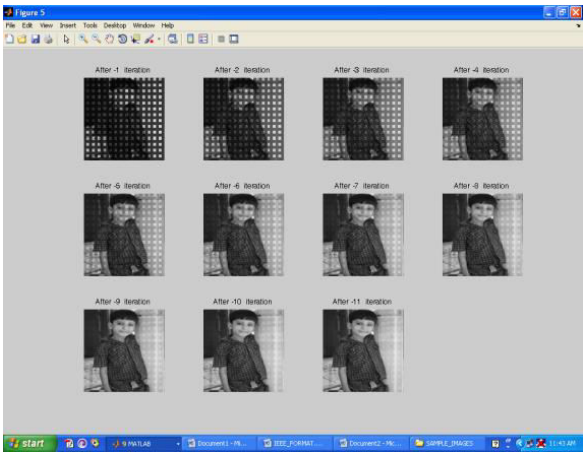
Compration betweens low resolution and high resolution image in terms of PSNR of Green sea turtle



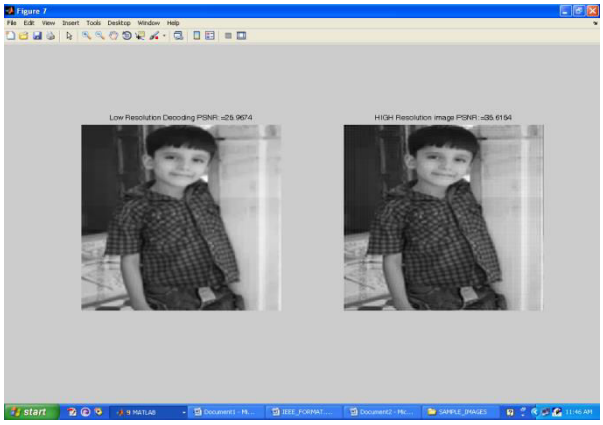
Frangipani flowers Decompressed Low Resolution image after 11 iteration



Compration betweens low resolution and high resolution image in terms of PSNR of Frangipani flowers



Boy image Decompressed Low Resolution image after 11 iteration



Comparison between low resolution and high resolution image in terms of PSNR of boy image

Where, $n \times m$ is the number of total pixels. X_{ij} and X'_{ij} are the pixel values in the original and reconstructed image respectively. The peak to peak signal to noise ratio (*PSNR* in dB) is calculated as

$$\text{PSNR} = 10 \log (255^2/\text{MSE})$$

Where usable grey level values range from 0 to 255.

$$\text{MSE} = (1/MN) \sum_{y=1}^{M} \sum_{x=1}^{N} [I(x,y) - I'(x.y)]^2$$

Where the sum over j, k denotes the sum over all pixels in the image and N is the number of pixels in each image.

Analysis :- From my complete analysis I found that compression ratio is directly portioned to peak signal to noise ratio from my technique not only improves the PSNR or the quality of the decompressed image, at the same time it retains the actual information.

VI. CONCLUSION & FUTURE SCOPE

There have been several techniques for image compression like the most widely adopted JPEG, SPIHT and so on. But the emergence of medical imaging and explode of images in this field has resurfaced the need to tailor made medical image compression techniques that can attain very high compression ratio, without compromising the medical information stored in the images. In this work we developed a two level compression technique. In the first level the image is compressed by rescaling the image to lower resolution, followed by transforming the image with lifting integer wavelets and encoded using SPIHT. At the decoder side, the decompressed image is scaled back to the original resolution using Papoulis-Gerchberg method of image interpolation. Results show a significant performance improvement over JPEG compression. But there are certain performance issues that need to be addressed.

Use of the proposed technique introduces redundant colours in the edge of ROI and non ROI region when the compression ratio is high (Such redundancy may lead to misleading diagnosis of the image in automated image processing technique. A future work of the proposed technique will be to remove the extra shade, that appears in the boundary of ROI part. This could be attained using either fuzzy based techniques or by adopting image inpainting techniques.

ACKNOWLEDGMENT

All acclamation to God, who has empowered and enabled us to accomplish the task successfully. First of all we would like to thank our God Almighty who really helps us in every problem during the project. We would like to express our sincere and humble gratitude to Almighty whose blessings, help and guidance has been a real source of all our achievements in our life.

REFERENCES

[1] M.A. Ansari ,R.S. Anand, Recent Trends In Image Compression And Its Application In Telemedicine And Teleconsultation, Xxxii National Systems Conference, Nsc 2008, December 17-19, 2008

[2] Debin Zhao, Y. K. Chan, And Wen Gao, Low-Complexity And Low-Memory Entropy Coder For Image Compression, Ieee Transactions On Circuits And Systems For Video Technology, Vol. 11, No. 10, October 2001

[3] Ahmed Abu-Hajar And Ravi Sankar, Region Of Interest Coding Using Partial-Spiht, Icassp, 2004, Ieee 2004

[4] Marco Cagnazzo, Sara Parrilli, Giovanni Poggi, And Luisa Verdoliva, Costs And Advantages Of Object-Based Image Coding With

[5] Shape-Adaptivewavelet Transform, Hindawi Publishing Corporation

[6] Eurasip Journal On Image And Video Processing Volume 2007, Article Id 78323, 13 Pages,Doi:10.1155/2007/78323

[7] Lin Ma, Feng Wu, Debin Zhao, Wen Gao, Siwei Ma, Learning-Based Image Restoration For Compressed Image Through Neighboring Embedding

[8] David Gibson1, Michael Spann And Sandra I Woolley, Diagnostically Lossless 3d Wavelet Compression For Digital Angiogram Video

[9] Emmanuel Christophe, W. A. Pearlman, Three-Dimensional Spiht Coding Of Hyperspectral Images With Random Access And Resolution Scalability, Proceedings Of The 40th Asilomar Conference On Signals, Systems, And Computers, October 29 - November 1, 2006, Pacific Grove, Ca, Usa

[10] D.Vijendra Babu, Dr.N.R.Alamelu, Wavelet Based Medical Image Compression Using Roi Ezw, International Journal Of Recent Trends In Engineering, Vol 1, No. 3, May 2009

[11] M.Tamilarasi, Dr.V.Palanisamy, Fuzzy Based Image Compression On ROI Using Optimized Directional Contourlet Transform, International Journal Of Recent Trends In Engineering, Vol 2, No. 5, November 2009

[12] Marcus Nyström, Jerry D. Gibson, John B. Anderson, Multiple Description Image Coding Using Regions Of Interest

[13] António J. R. Neves, Armando J. Pinho, Lossless Compression Of Microarray Images Using Image-Dependent Finite-Context Models, Ieee Transactions On Medical Imaging, Vol. 28, No. 2, February 2009

[14] Aysegul Cuhadar And Sinan Tasdoken, Multiple, Arbitrary Shape ROI Coding With Zerotree Based Wavelet Coders, Icassp 2003, Ieee-2003

◈◈◈

Face Verification Across Age Progression Using Discriminative Methods

Piyush D. Hegu[1], Ashwin W. Motikar[2] & Pradhnya J. Suryawanshi[3]

[1&2]Yavatmal (M.S.),India, [3]PCE, Nagpur, Nagpur(M.S.), India
E-mail : Piyush.hegu@gmail.com[1], ashwinmotikar@gmail.com[2], pradhnyajs@rediffmail.com[3]

Abstract - Face verification in the presence of age progression is an important problem that has not been widely addressed. This project is based on studying the problem by designing and evaluating discriminative approaches.First, find the gradient orientation (GO) which provides an effective representation. This representation is further improved when hierarchical information is used, which results in the use of the gradient orientation pyramid (GOP). When combined with a support vector machine (SVM), GOP demonstrates excellent performance in all our experiments

These experiments are conducted on the FGnet dataset and two large passport datasets, one of them being the largest ever reported for recognition tasks. Second, taking advantage of these datasets,we empirically study how age gaps and related issues (including image quality, spectacles, and facial hair) affect recognition algorithms.

Key words - *Face verification, age progression, gradient orientation pyramid, support vector machine.*

I. INTRODUCTION

A. Background

Face verification is an important problem in computer vision and has a very wide range of applications, such as surveillance, human computer interaction, image retrieval, etc.

Discriminative approaches have been used for face verification across age progression. The most related study to our work is [2], where the probabilistic eigenspace frame- work is adapted for face identification across age progression. Instead of using a whole face, only a half face (called a Point Five face) is used to alleviate the non-uniform illumination problem. Then, eigenspace techniques and a Bayesian model are combined to capture the intra-personal and extra-personal image differences. An Eigenspace is used in combination with a statistical model on the FGnet dataset [1] and on the MORPH dataset.

B. Tasks and challenges

The goal of our study is two-fold. The first is to investigate representations and algorithms for verification. The second is to study the effect of age gaps and related issues (including image quality, spectacles, and facial hair) on verification algorithms. We use three datasets in our study. Two of them are passport datasets involving more than 1,800 subjects, which to the best of our knowledge are the largest datasets ever studied for the task. We also use the FG-NET Aging Database [1] that is widely used for image based face aging analysis.

The challenges of face verification across age progression are due to several sources. The first source is the biometric change over years and the second source is the change in the image acquisition conditions and environment, including the illumination conditions. Some examples are as shown in figure below(Fig.1).

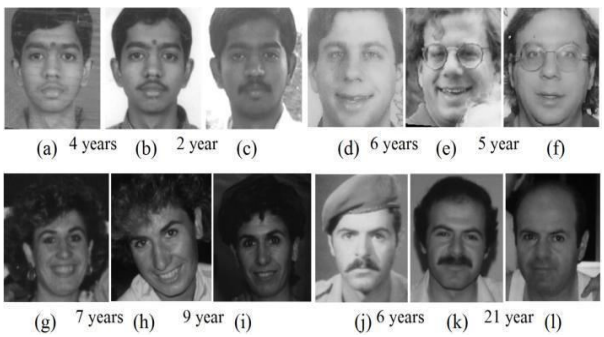


Fig. 1 : Typical images with age differences. Top row: scanned passport or visa photos. Bottom row: photos from the FG-NET Aging Database [1]

C. Contribution

We make several contributions in this study. First, we propose using the gradient orientation

pyramid (GOP) which is then combined with the support vector machine (SVM), GOP demonstrates excellent performance for face verification with age gaps.

Our second contribution is thorough empirical experiments. We evaluated nine different approaches, including two baseline methods ($l_2$norm and gradient orientation), four different representations with the same SVM-based framework (intensity difference, gradient with magnitude, gradient orientation, and GOP), the Bayesian face, and two commercial face verification systems.

Our third contribution is the empirical study of how verification performance varies with increasing age gaps and related issues.

II. PROBLEM FORMULATION

A. Face Verification Framework

In this paper, we study face verification tasks. In verification, one must determine whether two images come from the same person, as opposed to recognition, in which an individual is identified from a large gallery of individuals.

Given an input image pair $I_1$and I_2, the task is to assign the pair as either intra-personal (i.e. I_1 and $I_2$from the same people) or extra-personal (i.e. $I_1$and I_2 from different individuals). We use a support vector machine (SVM).

Note that gradient-based representations are recently widely used in computer vision and pattern recognition tasks, such as the scale invariant feature transfer (SIFT) for object and category classification and the histogram of orientation (HOG). In these works, the gradient directions were weighted by gradient magnitudes. In contrast, we discard magnitude information and Specifically, given an image pair (I_1, I_2), it is first mapped onto the feature space as

$$x = F(I_1, I_2) \qquad (1)$$

where $x \in Rd$ is the feature vector extracted from the image pair (I_1, I_2) through the feature extraction function $F : I \times I \rightarrow Rd$(F will be described in the following subsections),I is the set of all images, and R forms the d-dimensional feature space.

Then SVM is used to divide the feature space into two classes, one for intra-personal pairs and the other for extra- personal pairs. we denote the separating boundary with the following equation

$$\sum_{i=1}^{Ns} \alpha_i y_i K(s_i, x) + b = \Delta \qquad (2)$$

Where N_S is the number of support vectors and s_i is the i-th support vector. Δ is used to trade off the correct reject rate and correct accept rate as described in (3) and (4). K(., .) is the kernel function that provides SVM with non-linear abilities.In our experiments, we use the LibSVM library [8].

For verification tasks, the correct reject rate (CRR) and the correct acceptance rate (CAR) are two critical criteria,

$$CRR = \frac{\#\text{ correctly rejected extra-personal pairs}}{\#\text{ total extra-personal pairs}} \qquad (3)$$

$$CAR = \frac{\#\text{ correctly accepted intra-personal pairs}}{\#\text{ total extra-personal pairs}} \qquad (4)$$

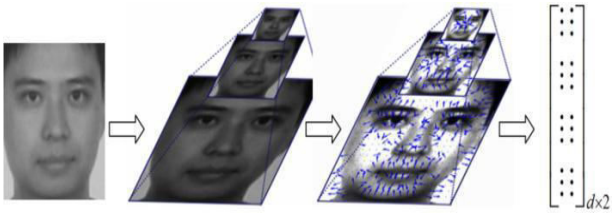
Where “accept” indicates that the input image pair are from the same subject and “reject” indicates the opposite.

B. Gradient Orientation and Gradient Orientation Pyramid

Now we need to decide the representation for feature extraction, i.e., *F(:; :)*. A natural choice is to use the intensity difference between *I*1 and *I*2, which is called *difference space*. The difference space can be made robust to affine lighting changes by an appropriate intensity normalization.

We propose to use GO for face verification across age progression. Specifically, in [3], GO is shown to be robust to illumination change and successfully applied for face recognition tasks. Furthermore, it has been shown in [6], [7] that the change of face color across age progression can be factored to two components, hemoglobin andmelanin, according to skin anatomy ,which shows that the GO of each color channel of human faces is robust under age progression. In addition, we collect gradient orientation in a hierarchical way, which has been shown to retain most visual information.

Note that gradient-based representations are recently widely used in computer vision and pattern recognition tasks , such as the scale invariant feature transfer (SIFT) for object and category classification and the histogram of orientation (HOG).



(a) Image I (b) Pyramid P (I) (c) GOP (d) G(I)

Fig. 2 : Computation of a GOP from an input image I. Note: In (c), the figure is made brighter for better illustration.

In these works, the gradient directions were weighted by gradient magnitudes. In contrast, we discard magnitude information and use only orientations, which demonstrates significant improvement in our experiments. Furthermore, the gradient directions at different scales are combined to make a hierarchical representation.

Given an image $I(p)$, where $p = (x; y)$ indicates pixel locations, we first define the pyramid of I as $P(I) = \{I(p; \sigma)\}^{S}_{\sigma=0}$ with:

$$I(p; 0) = I(p),$$
$$I(p; \sigma) = [I(p; \sigma - 1) * \Phi(p)] \downarrow 2 \quad \sigma = 1, ..., s, \quad (5)$$

where $\Phi(p)$ is the Gaussian kernel (0.5 is used as the standard deviation in our experiments), $\otimes$ denotes the convolution operator, $\downarrow 2$ operator, denotes half size downsampling, and s is the number pyramid layers.

Then, the gradient orientation at each scale σ is defined by its normalized gradient vectors at each pixel

$$g(I(p;\sigma)) = \begin{cases} \frac{\nabla(I(p,\sigma))}{|\nabla(I(p,\sigma))|} & \text{if } |\nabla(I(p,\sigma))| > \tau \\ (0,0)^{\top} & \text{otherwise} \end{cases}, \quad (6)$$

where τ is a threshold for dealing with "flat" pixels. The *gradient orientation pyramid* (GOP) of I, is naturally defined as $G(I) = \text{stack}(\{g(I(p, \sigma))\}_{\sigma=0} \in R \times 2$ that maps I to a $d \times 2$ representation, where stack(·) is used for stacking gradient orientations of all pixels across all scales and d is the total number of pixels. Fig. 2 illustrates the computation of a GOP from an input image.

C. *Kernels Between GOPs*

Given an image pair ($I1$; $I2$) and corresponding GOPs ($G1 = G(I1)$;$G2 = G(I2)$), the feature vector $x = F(I1; I2)$ is computed as the cosines of the difference between gradient orientations at all pixels over scales.

$$x = F(I_1, I_2) = (G_1 \odot G_2) \quad (7)$$

where $\odot$ is the element-wise product. Next, we apply the kernel to the extracted feature x to be used with the SVM framework. Specifically, our kernel is defined as

$$K(x1; x2) = \exp(-\gamma|x1 - x2|2) ; \quad (8)$$

Where γ is a parameter determining the size of RBF kernels ($\gamma = d$ is used in our experiments). In the rest of the paper, we use SVM+GOP to indicate the proposed approach.

III. FACE VERIFICATION EXPERIMENTS

A. *Experimental Setup*

Datasets : We conduct face verification experiments on three databases: two passport databases, named Passport I and Pass- port II, and the FGnet database [1]. All datasets are dominated by Caucasian descendants. Details of these databases are given in the following subsections. For computational reasons, image sizes are reduced to 96 × 84 for Passport I, 72 × 63 for Passport II, and 96 × 84 for the FGnet database.

Approaches : We compared the following approaches. 1) SVM+GOP: the approach proposed in this paper. 2) SVM+GO: this is similar to SVM+GOP, except that only the gradient orientation (GO) at the finest scale is used without a hierarchical representation. 3) SVM+G: this one is similar to SVM+GO, except that the gradient (G) itself is used instead of gradient orientation. 4) SVM+diff . We use the differences of normalized images as input features combined with SVM. 5) GO: this is the method using gradient orientation. 6) $l2$: this is a baseline approach that uses the $l2$ norm to compare two normalized images. 7) Bayesian+PFF. This is the approach combining Bayesian framework and PointFive Face (PFF).

The performance of algorithms is evaluated using the CRR-CAR curves that are usually created by varying some classifier parameters. We used threefold cross validation in our experiments. For each experiment, the CRR-CAR curve is created by adjusting parameter Δ. The total performance is evaluated as the average of the output CRR-CAR curves of three folds.

B. *Experiments with Passport Datasets*

We tested the proposed approach on two real passport image datasets, which we will refer to as Passport I and Passport II respectively. Passport I contains 452 intra-personal image pairs and 2,251 randomly generated extra-personal image pairs. Passport II contains 1,824 intra-personal image pairs and 9,492 randomly generated extra-personal image pairs. The extra-personal pairs are generated in the way such that there is no overlapping of subjects between training and testing sets. Images in both datasets are scanned passport images. They are in general frontal images with small pose variations. The lighting condition varies, and can be non-uniform and saturated. The age differences between image pairs are summarized in Table I. It shows that both datasets have significant age gaps for intra- personal images. Fig. 3 further shows the distribution of age differences of intra-personal pairs in the datasets whereas fig. 4 and Fig. 5 show the CRR-CAR curves for the experiments.

TABLE I

PASSPORT DATASETS FOR FACE VERIFICATION TASKS. "STD." IS SHORT FOR STANDARD DEVIATION.

Dataset	≠intrapair	Mean age	Std. age	Mean age diff.	Std. age diff.
Pass I	452	39	10	4.27	2.9
Pass II	1842	48	14.7	7.45	3.2

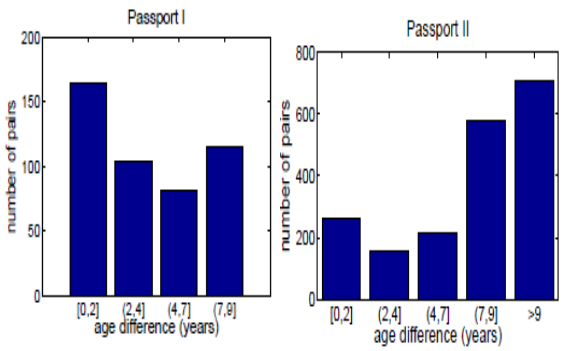


Fig. 3 : Distribution of age differences in the passport image databases. Left: Passport I, Right: Passport II.

There are several observations from the experimental results.

1) First, among the SVM-based approaches, GOP works the best. The gradient direction obviously plays a main role in GOP's excellent performance, since both SVM+GOP and SVM+GO largely outperform SVM+G, which includes the gradient magnitude information. In comparison, the use of a hierarchical structure in GOP further improves upon GO.

2) Second, SVM+GO greatly outperforms GO. Note that, for face verification, SVM+diff is previously used in [5] and GO is previously used in [3]. This shows that our method, as a combination of these two, greatly improves both of them.

3) Third, SVM+GOP outperforms the Bayesian approach on both datasets. In addition, from Fig. 5 it is obvious that SVM+GOP is more suitable for passport verification tasks because it performs much better at a high correct reject rate, which is desired as mentioned in Sec. II-A. Furthermore, given an image pair, our approach does not require the information of which one is older, which is used in the Bayesian approach as a prior.

4) Fourth, on Passport I,SVM+GOP performs similarly to Vendor A while much better than Vendor B, while on Passport II, SVM+GOP outperforms Vendor A but performs worse than Vendor B

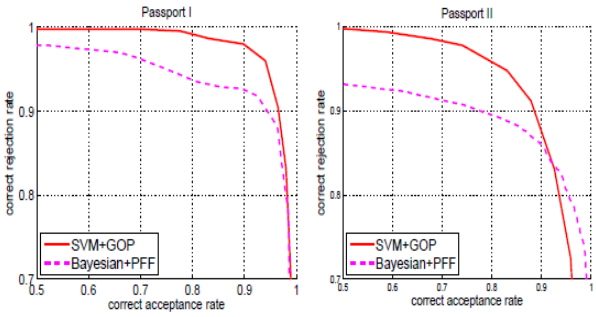


Fig. 5 : CRR-CAR curves for experiments with 200 intra- and 200 extra-pairs for training.

C. *Experiments on the FGnet Database*

The FGnet Aging Database [1] is widely used for research of age related facial image analysis. The database contains 1002 images from 82 subjects, over large age ranges. Consequently, there is an average of 12 images per subject in the FGnet database, which is much more than that in the passport databases (only two images per subject).

We use a subset of the FGnet database that contains only images that are taken above age 18 (including 18) and roughly frontal, which is consistent with the study on the passport databases. The effects of aging in children are quite different, and we discuss them in Section V. For notational

convenience, we still call this subset FGnet in the following

The subset contains 272 images from 62 subjects. Age statistics of FGnet are shown in Table III and Fig. 6.

We emphasize the importance of experiments on FGnet due to the following reasons:

- FGnet is very challenging for our task in two ways. First, it contains much larger age gaps. The largest gap is 45 years in FGnet, compared to 12 years in the passport databases. Second, the number of subjects is very limited, which makes learning very difficult.
- Since FGnet is a publicly available dataset, experiments on FGnet will serve as a benchmark/baseline for future studies on the topic.

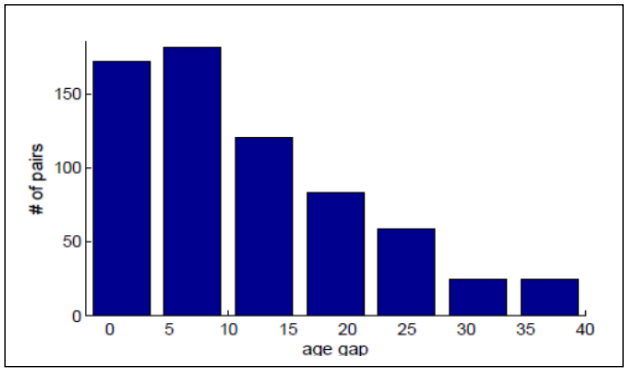


Fig. 6 : Distribution of age differences in the FGnet dataset.

TABLE II
FGNET DATABASE USED IN FACE VERIFICATION TASKS. "STD." IS SHORT FOR STANDARD DEVIATION.

#subject	≠intrapair	Mean age	Std. age	Mean age diff.	Std. age diff.
62	665	29.5	11.3	12.3	9.7

For verification tasks, we generate 665 intra-personal pairs by collecting all image pairs from same subjects. Extra- personal pairs are randomly selected from images from different subjects. Three-fold cross validation is used, such that in each fold images from the same subject never appear in both training and testing pairs. Each fold contains about 220 intra-personal pairs and 2,000 extra-personal pairs.

This confirms to some degree that our method is insensitive to illumination change, because PointFive Face is designed to be robust to illumination variations.

TABLE III

EQUAL ERROR RATES FOR EXPERIMENTS ON THE FGNET DATABASE

	L_2	GO	SVM +diff	SVM +G	SVM +GO	SVM+ GOP
EER	40.6%	32.3#	31.2#	28.5%	25.2%	24.1%

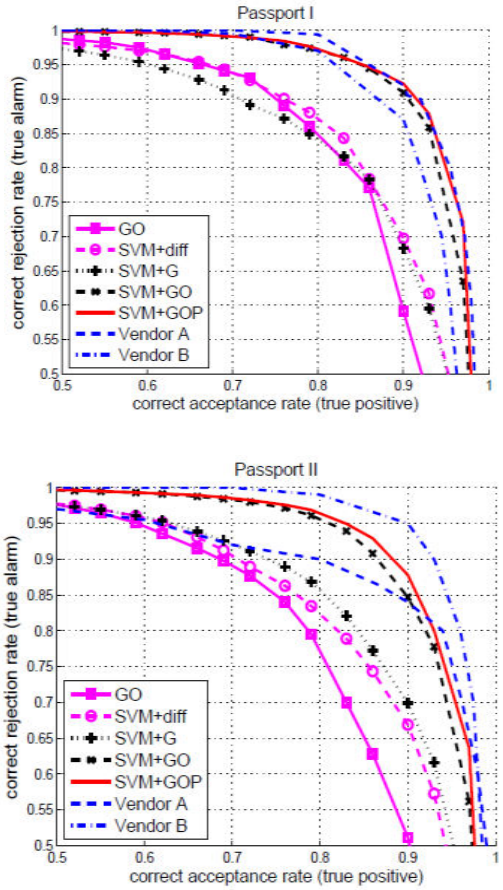


Fig. 4 : CRR-CAR curves for three-fold cross validation experiments. Top: on Passport I. Bottom: on Passport II.

TABLE IV

EQUAL ERROR RATES. UPPER TABLE: EXPERIMENTS OF THREE- FOLD CROSS VALIDATION LOWER TABLE: EXPERIMENTS USING 200 INTRA- AND 200 EXTRA-PAIRS AS TRAINING, AS IN [30].

	GO	SVM+ diff	SVM+G	SVM+ GO	SVM+ GOP
Pass I	17.6%	16.5%	17.8%	9.5%	8.9%
Pass II	20.7%	18.8%	17.4%	12.0%	11.2%

SVM+GOP	Bayesian
5.1%	8.5%
10.8%	12.5%

IV. EFFECTS OF AGE PROGRESSION ON VERIFICATION PERFORMANCE

In this section we empirically study how verification performance is affected by age gaps and related issues, including image quality, presence of eye glasses, and facial hair.

A. Effects of Age Gaps

We are interested in how age differences affect the performance of machine verification algorithms. Taking advantage of the large number of image pairs in Passport II.

First, intra-personal image pairs are grouped into age gaps from 0 to 2 years, 3 to 5 years, 6 to 8 years, and 9 to 11 years. The goal is to test verification performance for different groups. Specifically, we use the average equal error rates as a criterion. For each group, 80 intra pairs and 80 extra pairs are randomly selected as the training set. Testing sets are created similarly but with 15 intra pairs and 15 extra pairs. There is no overlap between training and testing sets. After that, four SVM-based approaches are tested on the data sets and equal error rates are recorded. To reduce the variance caused by the lack of training samples, 20 different training/testing sets are generated and the average equal error rates are recorded. The above experiments have been run 50 times with randomly chosen training/testing sets (i.e., 50 £ 20 training/testing sets). Finally, the mean and standard deviation of equal error rates are summarized to evaluate the performance.

Fig. 9 shows the performance of the experiments on all four groups. From the plots, we found that faces separated by more than a year are more difficult than those within one year. What surprised us is that the difficulty becomes saturated after the age gap is larger than four years. This phenomenon is observed on all four different representations tested in the experiments

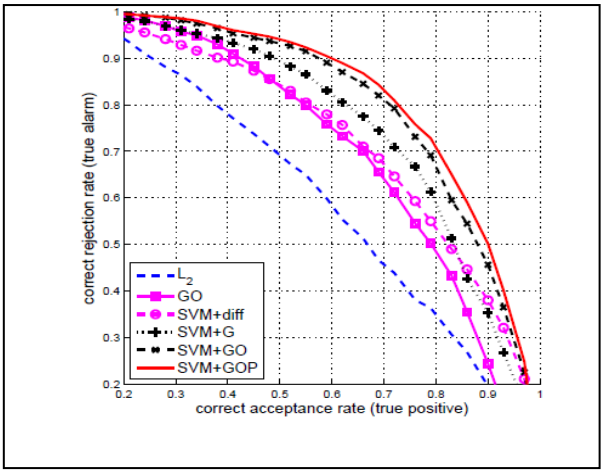


Fig. 7 : CRR-CAR curves for three-fold cross validation experiments on FGnet dataset. This figure is better viewed in color.

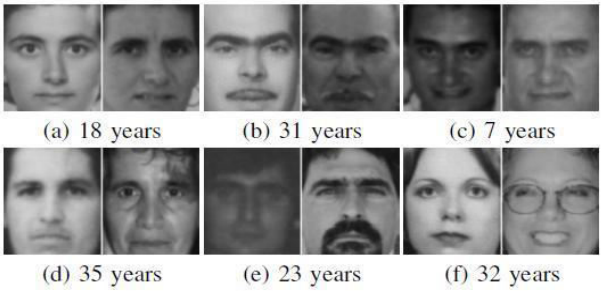


Fig. 8 : Example results of SVM+GOP on the FGnet datasets at the equal error rate. (a-c) Three correctly accepted intra- personal pairs. (d-f) Three incorrectly rejected intra-personal pairs. The listed years indicate age gaps in the corresponding pairs.

B. Effects of Age Related Issues

When comparing two images of the same person taken at different years, several non-anatomic issues often happen in practice. The FGnet dataset has detailed descriptions associated with each image. Using these descriptions, we analyze the verification results on the FGnet dataset to study the effects of the following three issues: 1) Quality, photos taken a long time ago sometimes have poor quality due either to the photographic environment or scanning artifacts. An intrapersonal pair is treated as high quality if both photos have good image quality and low otherwise. 2) Glasses: an intrapersonal pair is treated as different if one photo has spectacles and the other does not. Otherwise, the pair is treated as same. 3) Facialhair: an intra- personal pair is treated as without facial hair if none of photos has facial hair (including mustache and beard). Otherwise, the pair is treated as with facial hair. Once we have assigned each intra-pair with the

above labels, we can compare the error verification rate for each label and then compare how related issues affect verification algorithms. For example, the error rate of high (quality) inner-pairs is calculated as

$$1 - \frac{\#\text{ correctly classified high quality intra-pairs}}{\#\text{ high quality intra-pairs}}$$

Fig. 10 shows the error rates of different labels. These error rates are computed using SVM+GOP on the FGnet dataset and taken at the equal error rates (see Section III). From the figure, we see that low quality and spectacles do increase the difficulties for face verification. However, the proposed SVM+GOP seems to be robust to the presence of facial hair. One reason to this observation is, though facial hair sometimes adds difficulties to verification tasks, they often provide discriminative cues as well. For example, some people have similar beard styles over the years.

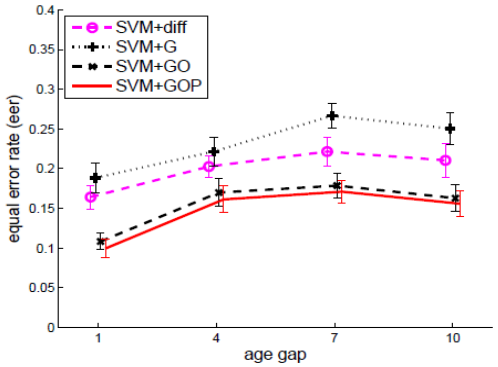


Fig. 9 : Effect of aging on verification performance. The curves are shifted a bit along the x axis for better illustration.

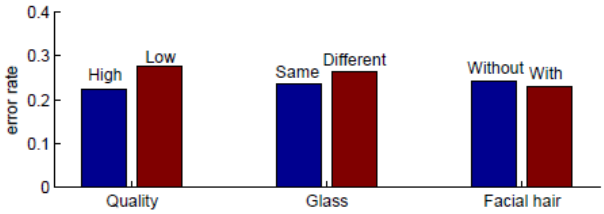


Fig. 10 : Error analysis of face verification experiments on the FGnet dataset.

V. FACE VERIFICATION ACROSS AGING IN CHILDREN

The appearance changes of human faces are very different in children than in adults [4]. In this paper we mainly focus on face images taken above age 18, after which face profiles remain stable [4]. However, it is helpful to understand the performance of the above tested methods on faces from children as well. In this section, we report our experiments on the children face images from the FGnet dataset.

We first extract two face datasets from FGnet, in the same way as in Sec. III-C. One dataset, named FGnet-18, contains 311 face images from 79 subjects, taken at ages in the range [8 18]. The other dataset, named FGnet-8, contains 290 face images from 74 subjects, taken at ages in the range [0 8].

For verification tasks, we follow the same scheme as in Sec. III-C; we generate 577 intra-personal pairs and 6,000 extra-personal pairs for FGnet-18, and 580 intra-personal pairs and 6,000 extra-personal pairs for FGnet-8. Three-fold cross validations are conducted for each dataset. Then, the average EERs and CRR-CAR curves are reported in Table V and Fig.11.

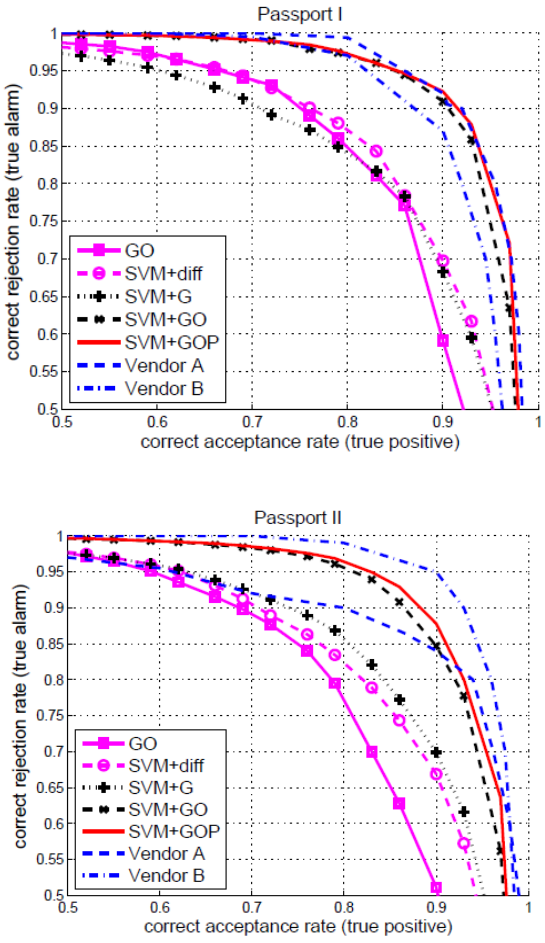


Fig. 11 : CRR-CAR curves for three-fold cross validation experiments on the children images of the FGnet dataset.

TABLE V

EQUAL ERROR RATES FOR EXPERIMENTS ON THE CHILDREN IMAGES OF FGNET DATABASE [1].

	L_2	GO	SVM +diff	SVM +G	SVM +GO	SVM+ GOP
FGnet -18	42.9 %	40.9 %	32.3%	36.1%	30.7%	30.5%
FGnet -8	44.0 %	44.6 %	36.2%	40.0%	39.8%	38.6%

From these experiments, we have the following observations. First, the verification tasks for childrens' faces are much harder than for adult faces. This is clear when we compare results in Table V and Table IV.

Second, gradient orientation based methods still work well for age changes of teenagers, though the hierarchical information does not help much any more. Third, the task becomes extremely difficult for small children with ages from 0 to 8, where all methods work poorly.

The major challenge of verifying children faces across aging comes from the alignment problem, because face profiles undergo large variations before age 18. This explains why the intensity (after normalization) based method, SVM+diff, works relatively better. Generative approaches can provide helpful guidance here, though age information is often requested. It is an interesting future direction to combine generative and discriminative approaches for this task.

VI. CONCLUSION AND DISCUSSION

In this paper we studied the problem of face verification with age variation using discriminative methods. First, we proposed a robust face descriptor, the gradient orientation pyramid, for face verification tasks across ages. Compared to previously used descriptors such as image intensity, the new descriptor is more robust and performs well on face images with large age differences. In our experiments with comparison to several techniques, the new approach demonstrated very promising results on two challenging passport databases and the FGnet dataset. In addition, being a discriminative approach, the proposed method requires no prior age knowledge and does not rely on age estimation and simulation algorithms.

Second, the effect of the aging process on verification algorithms are studied empirically. In the experiments we observed that the difficulty of face verification algorithms saturated after the age gap is larger than four years (up to ten years). We also studied the effects of age related issues including image quality, presence of spectacles, and facial hair.

REFERENCES

[1] "Face and gesture recognition working group," 2000.[Online]. Available: http://www-prima.inrialpes.fr/FGnet/

[2] N. Ramanathan and R. Chellappa, "Face verification across age progression,"IEEE Transactions on Image Processing, vol. 15, no. 11, pp. 3349–3361, 2006

[3] H. Chen, P. Belhumeur, and D. Jacobs, "In search of illumination invariants," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, 2000, pp. 254–261.

[4] J. B. Pittenger and R. E. Shaw, "Aging faces as viscal-elastic events: Implications for a theory of nonrigid shape perception," J. of Exp. Psychology: Human Perception and Performance, vol. 1, no. 4, pp. 374–382, 1975.

[5] P. J. Phillips, "Support vector machines applied to face recognition," in Advances in Neural Information Processing Systems 16 (NIPS), vol. 2, 1999, pp. 803–809.

[6] N. Tsumura, H. Haneishi, and Y. Miyake, "Independent component analysis of skin color image," Journal of Optical Society of America A, vol. 16, pp. 2169–2176, 1999.

[7] N. Tsumura, N. Ojima, K. Sato, M. Shiraishi, H. Shimizu, H. Nabeshima, S. Akazaki, K. Hori, and Y. Miyake, "Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin." ACM Trans. Graph., vol. 22, no. 3, pp. 770–779, 2003.

[8] C.-C. Chang and C.-J. Lin, "Libsvm : a library for support vector machines," 2001. [Online]. Available: http://www.csie.ntu.edu.tw/ cjlin/libsvm

◈◈◈

Dynamic Stability Improvement With Combined Fast Acting Automatic Voltage Regulator (AVR) And Automatic Load Frequency Control (ALFC) Loops

Venu Yarlagadda[1] , K. R. M. Rao[2], P. Sai Saroja[3], P. Avinash[4] & P. Nikhilesh[5]

[1,3,4&5]VNR VJIET , [2] MJCET

E-mail : venuyar@gmail.com[1], drkrmrao@yahoo.com[2]

Abstract - The dynamic stability of the Synchronous Generator have been improved using fast acting combined Automatic Voltage Regulator (AVR) and Automatic Load Frequency Control (ALFC) loops. This paper investigates the effect of Fast acting AVR and ALFC loops on the dynamic stability improvement.

The combined linearised AVR and ALFC model has been simulated in Matlab/Simulink with medium time constants (case 1) and low values (faster) of time constants (case 2). The block diagram of Automatic load frequency control (ALFC) of an isolated power system (obtained by combining the individual blocks of generator, load, prime mover model and speed governing system) and AVR system with PID controller. The frequency deviation step response and terminal voltage step response have been obtained for both the cases. The paper compares the performance of both medium and fast acting loops on dynamic stability improvement

Key words *- AVR Control, ALFC Control, Combined AVR and ALFC loops, Dynamic Stability, Stability Improvement using Fast acting controllers and Steam Power Station Stability Analysis.*

I. INTRODUCTION

In this paper the dynamic stability analysis of a steam power station using Matlab/Simulink have been studied. The combined fast acting Automatic load frequency control (ALFC) and Automatic voltage regulator (AVR) control loops have been simulated in SIMULINK. The load frequency loop controls the real power and frequency and the automatic voltage regulator loop regulates the reactive power and voltage magnitude. The excitation system time constant is much smaller than prime mover time constant and it's transients decay much faster and does not effect the ALFC dynamics. The cross coupling between the LFC loop and the AVR loop is negligible and the load frequency and excitation voltage control are analysed independently and its simulation responses are obtained.

The MATLAB/SIMULINK Models have been developed to study the dynamic stability analysis of the system. The simulation response of individual ALFC, AVR and combined ALFC and AVR loops has been studied. From the simulation responses of these systems the information about steady state frequency deviation step response and time taken for the frequency to return to its nominal value is obtained for the different values of time constants. The results show the effect of fast acting Combined AVR and ALFC loops on Dynamic Stability.

II. CLASSIFICATION OF STABILITY

The stability of an interconnected power system is its ability to return to its normal or stable operation after having been subjected to some form of disturbance. The tendency of synchronous machine to develop forces so as to maintain synchronism and equilibrium is called stability. The stability limit represents the maximum stead state power flow possible when the synchronous machine is operating with stability.

There are three forms of stability.

- Steady state stability
- Transient stability
- Dynamic Stability

Stability Improvement

Methods to improve stability are:

1. Use of double circuit lines
2. Use of bundle conductors.
3. Series compensation of the lines

4. High speed excitation systems
5. HVDC links
6. Load shedding
7. Using higher excitation voltage
8. Using the quick response system

III. LOAD FREQUENCY CONTROL OF AN ISOLATED POWER SYSTEM

The operating objectives of the Load Frequency Control (LFC) are to maintain reasonably uniform frequency to divide the load between generators, and to control tie-line interchange schedules. The change in frequency and tie-line power are sensed, which is a measure of the change in rotor angle δ, i.e the error Δδ to be corrected. The error signal, i.e., Δf and ΔP_{tie}, are amplified, mixed, and transformed into a real power signal ΔP, which is sent to the prime mover to call for an increment in the torque.

The prime mover therefore brings change in the generator output by an output ΔP_g which will change the values of Δf and ΔP_{tie} with the specified tolerance.

The system load changes continuously, the generation are adjusted automatically to restore the frequency to the nominal value. The scheme is known as the *automatic generation control*. In an inter connected system consisting of several loops, the role of AGC is to divide the loads among system, station, and generators so as to achieve maximum economy and correctly control the scheduled interchanges of tie-line power while maintaining a reasonably uniform frequency.

With the primary LFC loop, a change in the system load will result in a steady-state frequency deviation depending on the governor speed Regulation. In order to reduce the frequency deviation to 0, a reset action is provided by introducing an integral controller to act on the load reference setting to change the speed set point. The Block Diagram Model of ALFC Loop simulated in MATLAB/SIMULINK as shown in fig.1 below.

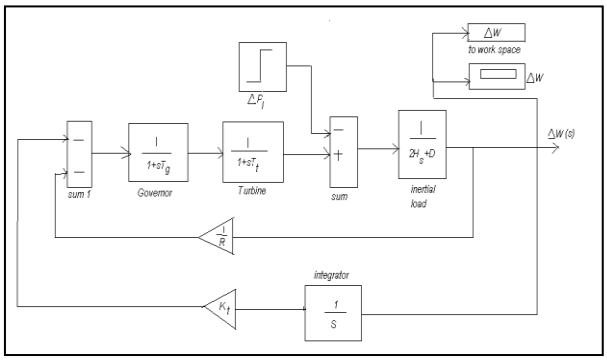


Fig. 1: Block Diagram Model of ALFC Loop

IV. AUTOMATIC VOLTAGE REGULATOR

A change in the real power demand effects essentially the frequency, whereas a change in the reactive power affects mainly the voltage magnitude. The interaction between voltage and frequency controls is generally weak enough to justify their analysis separately. The sources of reactive power are generators, capacitors, Synchronous Condensers and FACTS Controllers. The generator reactive power is controlled by field excitation. The primary means of generator reactive power control is generator excitation control using *automatic voltage regulator* (AVR) is to hold the terminal voltage magnitude of a synchronous generator at desired level. The Block Diagram Model of AVR Loop simulated in MATLAB/SIMULINK as shown in fig.2 below.

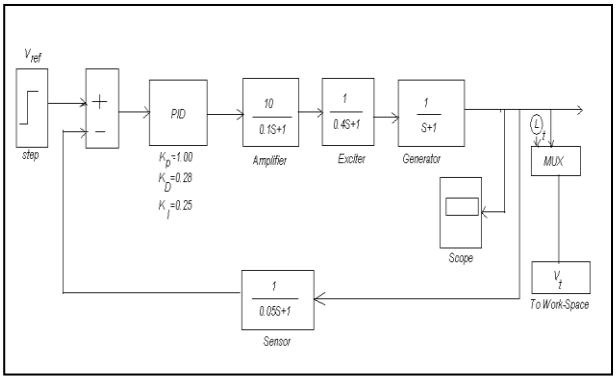


Fig. 2 : Block Diagram Model of AVR Loop

V. LINEARISED MODEL OF COMBINED ALFC AND AVR LOOPS

In an interconnected power system, Automatic load frequency control (ALFC) and Automatic Voltage Regulator (AVR) equipment is installed for each generator. The controllers are set for a particular operating condition and take care of small changes in load demand to maintain the frequency and voltage magnitude within the specified limits. The excitation system time constant is much smaller than the prime mover time constant and its transient decay much faster than and does not affect the LFC dynamic. Thus the cross coupling between the LFC loop and the AVR loop is negligible, and the load frequency and excitation voltage control are analysed interpedently. The schematic diagram of Combined ALFC and AVR loops as shown in fig.3 below and the combined MATLAB/SIMULINK model as shown in fig.4 below.

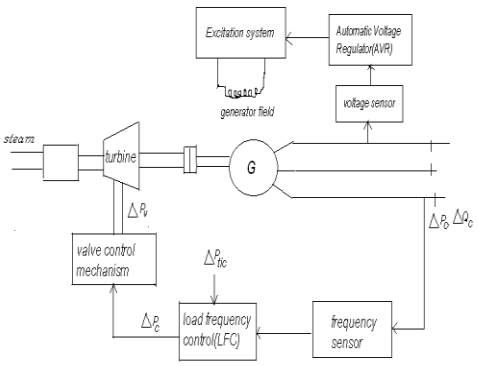


Fig. 3: schematic diagram of Combined ALFC and AVR loops

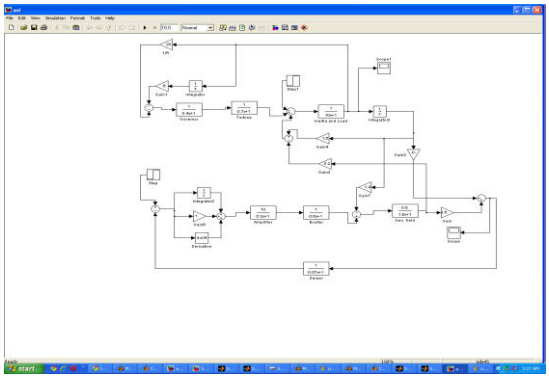


Fig. 4: Block Diagram of combined AVR and ALFC Loops

Case 1: Linearised Model of Combined ALFC and AVR Loop Parameters

Table 1: Case 1 gain and time constants of Combined system (medium values)

S.No	Component	Gain	Time Constant
1	Governor	$K_{g1}=1$	$T_{g1}=0.4$
2	Turbine	$K_{t1}=1$	$T_{t1}=0.7$
3	Amplifier	$K_{a1}=10$	$T_{a1}=0.3$
4	Exciter	$K_{e1}=1$	$T_{e1}=0.6$
5	Generator	$K_{g1}=0.8$	$T_{g1}=1.6$
6	Sensor	$K_{r1}=1$	$T_{r1}=0.07$
7	Inertia		$H_1=5$
8	Regulation		$R_1=0.05$
9	Proportion controller		$K_{p1}=1$
10	Integral controller		$K_{t1}=0.25$
11	Deviation Controller		$K_{d1}=0.3$

Case 2: Linearised Model of Combined ALFC and AVR Loop Parameters

Table 2: Case 2 gain and time constants of Combined system (low values)

S.No	Component	Gain	Time Constant
1	Governor	$K_{g2}=1$	$T_{g2}=0.2$
2	Turbine	$K_{t2}=1$	$T_{t2}=0.5$
3	Amplifier	$K_{a2}=10$	$T_{a2}=0.1$
4	Exciter	$K_{e2}=1$	$T_{e2}=0.4$
5	Generator	$K_{g2}=0.8$	$T_{g2}=1.4$
6	Sensor	$K_{r2}=1$	$T_{r2}=0.05$
7	Inertia		$H_2=5$
8	Regulation		$R_2=0.05$
9	Proportion controller		$K_{p2}=1$
10	Integral controller		$K_{t2}=0.25$
11	Deviation Controller		$K_{d2}=0.3$

VI. CASE STUDY

Case 1: Results

The per unit steady state frequency deviation is 0.2 and time taken for this is more than 10 sec as shown in fig.5 below.

Fig. 5: Frequency deviation step response for Case 1

The terminal voltage response of linearised model for the combined LFC and AVR system is satisfactory with a settling time of 10sec as shown in fig.6 below

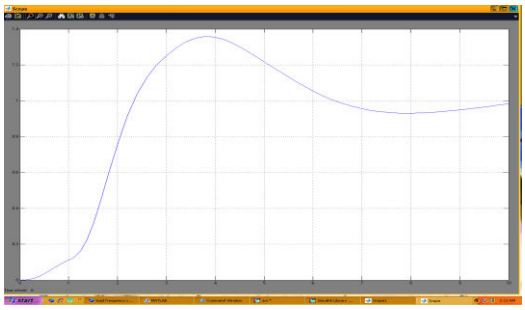


Fig. 6: Terminal voltage step response for Case 1

Case 2: Results

The per unit steady state frequency deviation is 0, approximately and time taken for the steady state frequency deviation is 10 sec as shown in fig.7 below.

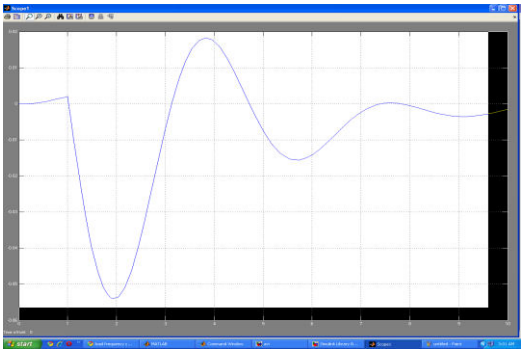


Fig. 7: Frequency deviation step response for Case 2

The response of Linearised Model for the combined LFC and AVR system is satisfactory with a settling time of 9 sec as shown in fig.8 below.

Fig. 8: Terminal voltage step response for Case 2

The above results shows the per unit frequency deviation step responses and the terminal voltage step responses for the Linearised model of the combined LFC and AVR system for medium, low value of time constants

VII. CONCLUSIONS

In this Paper, the Frequency deviation step response and Terminal voltage step responses have been obtained for the steam power station with individual ALFC, AVR and Linearized model of the combined ALFC and AVR loops.

In the first case (Case 1) the combined block diagram model has been simulated in MATLAB/SIMULINK with medium time constants. In the second case (Case 2) the same combined block diagram model has been simulated with faster loops (low values of time constants). The results shows that with fast acting combined ALFC and AVR loops improves the dynamic response of both frequency and voltage significantly.

Future Scope

The dynamic stability of a Power Station can be improved practically by choosing fast acting devices of ALFC and AVR loops and it can be physically realised for a Steam Power Station. Power System stabilizers (PSS) can also be used as a supplementary controller along with Fast acting AVR and ALFC loops for dynamic stability improvement.

ACKNOWLEDGEMENTS

We are immensely thankful to Management and Principal of VNR VJIET for providing R&D lab and other facilities to carryout this work.

NOMENCLATURE

AVR : Automatic Voltage Regulator

AGC : Automatic Governor Controller

ALFC : Automatic Load Frequency Control

LFC : Load Frequency Control

PID : Proportional Integral and Derivative

PSS : Power System Stabiliser

δ : load angle or power angle

$\Delta\delta$: incremental change in load angle

Δf : incremental change in frequency

ΔP_{tie} : incremental change in Tie-line power

ΔP_g: incremental change in Output Power

REFERENCES

[1] N.G.Hingorani and L. Gyugyi, Understanding FACTS, IEEE Press, New York, USA, 1999.

[2] Power System Dynamics Stability and Control, 2nd Edition - K R PADIYAR

[3] IEEE Power Engg. Society/CIGRE, "FACTS Overview", Publication 95 TP 108, IEEE Press, New York, 1995.

[4] P. Kundur, Power System Stability and Control, McGraw-Hill, Inc., New York, 1994.

[5] W.A.Lyon, Transient Analysis of Alternating Current Machinaery, Chapter 2, John Wiley, New York, 1954.

◈◈◈

Design And Implementation For Identification Of Moisture Content In Cotton Bale By Microwave Imaging

Ashish V. Saywan

Dept. of Electronics (Communication), P.I.E.T., Nagpur, India
E-mail : ashishvsaywan@gmail.com

Abstract - Understanding of cotton quality is important in order to properly identify the moisture content .Measurement of moisture is difficult particularly at harvest and through the gin, because of the influence these processes have different fibre quality. Dry cotton can be harvested cleanly and efficiently but may suffer undue damage in the gin. On the other hand harvesting and ginning wet cotton leads to significant issues in processing and quality. A number of methods are used to measure moisture in seed cotton, lint and fuzzy seed, each has its varying advantages. A moisture variation of the bales that is not monitored from the outside of the bale. This research examines a new microwave imaging technique to view the internal moisture variations of cotton bale. Tests on the developed imaging sensor showed the ability to resolve small structures of parameters, against a low standard background, that were less than 1 cm in width. The accuracy of the sensing structure was also shown to provide the ability to accurately determine parameter standards. A preliminary test of the imaging capabilities on a wet commercial bale showed the technique was able to accurately image and determines the location of the wet layer within the bale.

Key words - *Bales, fibre, gigging, moisture, microwave imaging.*

I. INTRODUCTION

In cotton grading the quality of cotton is identified by steple length, moisture level strength of cotton thread present all these parameter are identified manually, by the operator. In order to standardize this process the cotton project is intended. This project is focused on identification of moisture level in the cotton bale and generates appropriate result. The final stage in the cotton processing stream is the cotton bale packaging system. Recent innovations have shown that the use of cotton moisture restoration systems both reduce stress on the bale packaging system as well as add additional weight to the bales. As cotton is sold on a wet basis, these systems were beginning to proliferate through the ginning industry. The research is needed because the excess moisture in the bales will cause fiber degradation and color-grade changes. This issue has become so important that it prompted the Cotton industry to make a recommendation for bales to be limited not to be less than 7.5% moisture.

II. APPROACHES IN MEASURING MOISTURE

The methods for measuring moisture in cotton lint can be classified into six groups based on the technique and on the type of cotton material being tested, such as compacted or loose seed cotton, loose lint moved by air in ducting or compressed baled lint. Moisture measurement methods can be based on:

1. Thermal drying.
2. Chemical.
3. Spectroscopy.
4. Measurement of electrical or dielectric properties.
5. Compression properties of cotton lint.

III. WHAT IS MICROWAVE IMAGING METHOD?

This method involving the transmission of microwaves (electromagnetic radiation) signal of range between 2400-2480 MHz signal through the bale to determine change of signal in order to penetrate the these signal through a cotton bale without interrupting of signal to detect the change in phase and signal strength at receiver end.

IV. WHAT IS THE ROLE OF CC2500 MODULE?

C2500 is a Low-Cost Low-Power RF Transceiver. It functions in 2400- 2483.5 MHz frequency band and provides an excellent option for WSN applications because of its low-power characteristics and SRD (Short Range Device) frequency band.

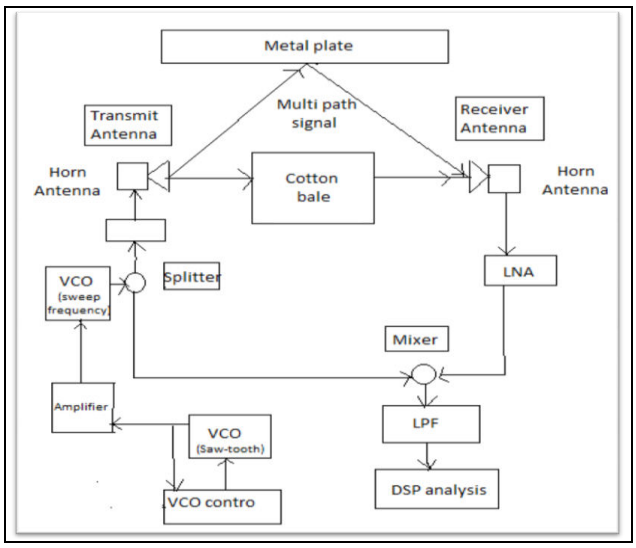


Fig. 1: Schematic layout of detection of moisture content in cotton bale using microwave method.

This chip has 20 pins as,

- 2 for connecting a (mandatory) 26MHz external Crystal oscillator.
- 2 for connecting the antenna.
- 10 for powering the chip.

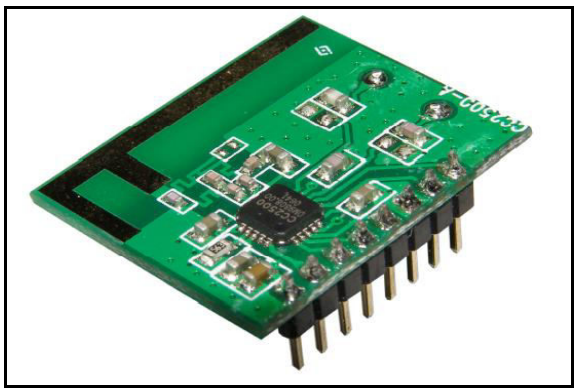


Fig.1 : CC2500 Transceiver module

- 6 for digital communication with the Microcontroller.
- The chip contains 47 registers to configure operating frequency, modulation scheme, baud rate, trans- mission power, etc. Because these registers are erased during power down, the MSP430 should configure all of them at startup.
- 13 commands allow the Microcontroller to control the state of the CC2500 (transmit, power Down, receive, etc).

The RF transceiver is integrated with a highly configurable baseband modem. The modem supports various modulation formats and has a configurable data rate of up to 500 k Baud. CC2500 provides extensive hardware support for packet handling, data buffering, burst transmissions, clear channel assessment, link quality indication, and wake-on-radio. The main operating parameters and the 64- byte transmit/receive FIFOs of CC2500 can be controlled via an SPI interface. In a typical system, the CC2500 will be used together with a microcontroller and a few additional passive components.

Pin No.	Name of Pin
1.	SCLK
2.	SO(GD01)
3.	S1(GD02)
4.	GVDD
5.	DCOUPLE
6.	GD00(ATEST)
7.	CSn
8.	XOSC_Q1
9.	AVDD
10.	XOSC_Q2

Fig Pin Configuration of CC2500

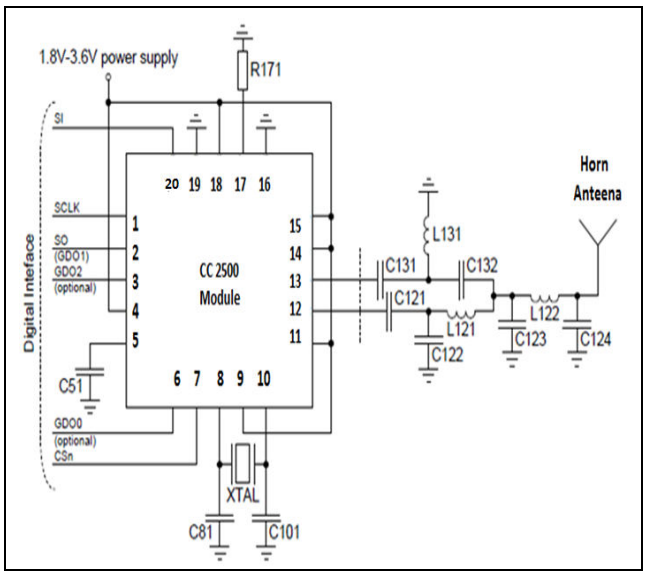


Fig2. Pin diagram of CC2500 module.

V. WHAT IS THE ROLE OF MSP430 MICROCONTROLLER?

MSP430 is a great fit for any mobile application, where power conservation is a priority. The many power-saving mechanisms designed into the MSP430 make it ideal for such applications. An emerging application is ISM-band (Industrial, Scientific, and Medical) and SRD-band (Short Range Device) wireless

connections, in the 315/433/868/915-MHz and 2.4-GHz bands. Markets served by this application include AMR (Automatic Meter Reading), low-power telemetry, and wireless sensor networks. It supports RF link with CC2500 forms a highly power-efficient wireless node that can transceive data at rates up to 500 kbps. The CC2500 is equipped with a SPI port, through which they can communicate with an MSP430

The Texas Instruments (TI) MSP430 family of processors are low power 16 bit devices. They are marketed at low power applications such as battery devices, however in spite of this they contain a sophisticated processor core and depending upon the model, a useful array of integrated peripherals. It produces a set of microcontrollers based around a core 16 bit CPU as shown in Figure.

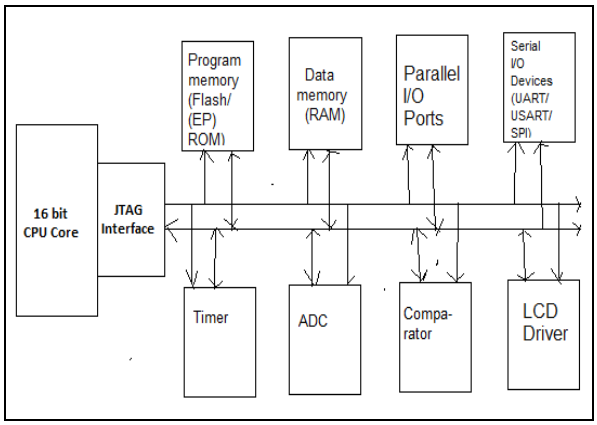


Fig. 3: Block diagram of MSP 430 uC

Some features of msp430:

- Ultra-low-power 16-bit RISC mixed-signal processors.
- Low power & high performance Seven source-addressing modes.
- Four destination-address modes.
- Only 27 core instructions.
- Prioritized, nested interrupts.
- No interrupt or subroutine level limits.
- Large register file.
- Ram execution capability.
- Efficient table processing.
- Fast hex-to-decimal conversion

What is SPI-Interface?

Most of the communication between the controller and the CC2500 is done using industry standard SPI(Serial Peripheral Interface)locks that work reliably with CC2500. However, the SPI-interface of the CC2500 requires a chip select -signal which must be generated on software. In addition to the SPI-interface, the CC2500 also provides two extra signals GD0 and GD2. Function of these signals can be selected rather freely (see datasheets of the CC2500 for details). In the most basic design these signals are used to signal the Controller when a valid packet has been received and transmit is complete. In this design, handling these signals is done completely using software. Among other options, these signals can also be used to transmit and receive asynchronous data up to 250 kbps.

SPI is implemented in the MCU by a hardware module called the Synchronous Serial Port or the Master Synchronous Serial Port SPI is a Data Exchange protocol. As data is being clocked out, new data is also being clocked in.

When one “transmits” data, the incoming data must be read before attempting to transmit again. If the incoming data is not read, then the data will be lost and the SPI module may become disabled as a result. Always read the data after a transfer has taken place, even if the data has no use in your application.

Data is always “exchanged” between devices. No device can just be a “transmitter” or just a “receiver” in SPI. However, each device has two data lines, one for input and one for output.

These data exchanges are controlled by the clock line, SCK, which is controlled by the master device.

RF Performance

- High sensitivity (–104 dBm at 2.4 kBaud, 1% packet error rate).
- Low current consumption (13.3 mA in RX, 250 kBaud, input well above sensitivity limit).
- Programmable output power up to +1 dBm.
- Excellent receiver selectivity and blocking performance.
- Programmable data rate from 1.2 to 500 kBaud.
- Frequency range: 2400 – 2483.5 MHz.

Analog Features

- OOK, 2-FSK, GFSK, and MSK supported.
- Suitable for frequency hopping and multichannel systems due to a fast settling frequency synthesizer with 90 us settling time.
- Automatic Frequency Compensation (AFC) can be used to align the frequency synthesizer to the received centre frequency.
- Integrated analog temperature sensor.

Digital Features

- Flexible support for packet oriented systems: On-chip support for sync word detection, address check, flexible packet length, and automatic CRC handling.
- Efficient SPI interface: All registers can be programmed with one "burst" transfer.
- Digital RSSI output Programmable channel filter bandwidth.
- Programmable Carrier Sense (CS) Indicator.
- Programmable Preamble Quality Indicator (PQI) for improved protection against false sync word detection in random noise.
- Support for automatic Clear Channel Assessment (CCA) before transmitting (for listen-before-talk systems).
- Support for per-package Link Quality Indication (LQI).
- Optional automatic whitening and de-whitening of data.

Low-Power Features

- 400 nA SLEEP mode current consumption
- Fast startup time: 240 us from SLEEP to RX or TX mode (measured on EM design)
- Wake-on-radio functionality for automatic low-power RX polling
- Separate 64-byte RX and TX data FIFOs (enables burst mode data transmission)

General

- Few external components: Complete on chip.
- Frequency synthesizer, no external filters or RF switch needed.
- Green package: RoHS compliant and no antimony or bromine
- Small size (4x4 mm package, 20 pins)
- Support for asynchronous and synchronous serial receive/transmit mode for backwards compatibility with existing radio communication protocols.

VI. CONCLUSION

This research is part of Quality and Utilization of Agricultural Products, The designing and implementation of this project is helping the cotton industry to design a particular criterion to purchase the moisture free cotton from the market and accordingly increasing the quality, accuracy and precision of the cotton for further processing.

REFERENCES

[1] Anthony, W.S. 2003a. Impact of Moisture on Baled Cotton. American Society Of Agricultural Engineers Meetings Papers. Paper No. 031167, 37pp.

[2] Anthony, W.S. 2003b. The Impact of Excess Moisture in The Bale On Fiber Quality. National Cotton Council Belt wide Cotton Conference.

[3] Baker, K.D., Hugh's, S.E., McAlister, D.D. 2005. Bale moisture addition with a rotor spray system. In: Proceedings of the Belt wide Cotton Conference, January 4-7, 2005, New Orleans, Louisiana. p. 808-812. 2005

[4] R.K. Byler and W.S. Anthony, 1995. Development of a moisture sensor for gin use. Proc. Belt wide Cotton Conf., San Antonio, TX. 4-7 Jan. 1995. Natl. Cotton Counc. Am., Memphis, TN.

[5] R.K. Byler and W.S. Anthony, 1997. Measurement Concepts in a gin process control system, Proc. Belt wide Cotton Conf., New Orleans, LA. 7-10 Jan. 1997. Natl. Cotton Counc. Am., Memphis, TN.

[6] Byler, R.K., Anthony, W.S. 2003. Measurement Of Bale Moisture Content Based On Bale Compression Pressure. 2003 Belt wide Cotton Conference.

[7] Mangialardi, G. J. Jr., and A. C. Griffin, Jr. 965. Moisture restoration to cotton at the gin: effects n fiber and spinning properties. USDA Report o. 08. Washington, D.C.: USDA.

[8] Mangialardi, G. J. Jr., and A. C. Griffin, Jr. 1977. A method for restoring moisture to cotton at gins. *Transactions of the ASAE* 20(5).

[9] Shepherd, J. V. 1972. Standard procedures for foreign matter and moisture analytical tests used in cotton ginning research. USDA Agricultural Handbook No. 422. Washington, D.C.: USDA.

[10] Pelletier, M.G. 2005b. Bale moisture sensing: Field Trial. 2005 Belt wide Cotton Conference.

◈◈◈

Application of Virtual Routers in Wireless Communication

Ashwin W. Motikar[1] & Narendra Narole[2]

[1]Dept. of Communication, Nagpur(M.S.), India
[2]Dept. of Electronics, Priyadarshini Inst. of Engg. & Tech., Nagpur(M.S.), India
E-mail : ashwinmotikar@gmail.com[1]

Abstract - The deployment of virtualized network resources has the potential to spur new business models and increase flexibility for network customers as well as infrastructure op- erators. It is worthwhile to re-evaluate how to effectively express traditional network elements in the virtualization domain. In this paper we consider network routers and argue that the representation of routing functionality as a service, rather than an isolated virtual resource is better suited in the virtualization context.

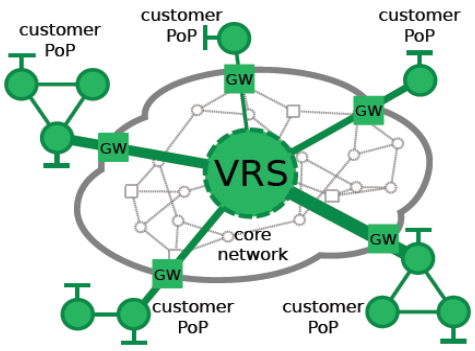
We present an architecture enabling physical infrastructure operators to provide routing as a service. To this end, distributed forwarding elements are combined to appear a single virtual router instance which routes traffic between a set of customer points of presence. We provide embedding algorithms for virtual router topologies with minimum allocation cost. We consider the customer's geographical attachment to the network, bandwidth demands as well as capacity constraints in the core substrate. Moreover, we present a live-migration approach for the virtual router data plane which allows network operators to quickly adapt resources to changing network demands.

I. INTRODUCTION

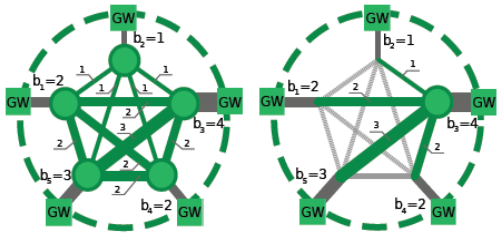
To date, a substantial amount of research in the network virtualization domain has focused on the embedding of pre- defined virtual network topologies onto a physical substrate, a problem known to be NP-hard. Heuristics for the general sub- strate embedding problem approximating the optimal solution have been proposed e.g. [1], [2], [3]

In this paper we advocate the concept of virtual routers as a service - a collection of virtual network resources functioning as a single router instance as illustrated in Fig.1a. We believe that routing functionality in virtual networks is more suitably defined in terms of connectivity between end points rather than topologies mimicking physical networks. Traditional design goals such as resilience are likely to remain a responsibility of the physical infrastructure provider, addressed independently of the virtual domain instantiation. An inherent advantage of this simplified viewpoint is that the substrate embedding problem becomes tractable. We discuss algorithms for the op- timal allocation of resources in capacity constrained substrate networks. In addition, we develop a flexible architecture for virtual router services (VRS).

VRS can be deployed to consolidate physical provider resources and adapt substrate allocation to changing network conditions without disrupting running services. At the same time, customers can reduce the number of physically hosted devices while seamlessly integrating their router instance into an existing infrastructure. Additional aspects of the single router abstraction as a means for facilitating network man- agement are discussed in the position paper [4].



(a) A virtual router service connecting five customer locations with a specific capacity demand over a provider substrate.



(b) Fully connected topology (c) Star topology

Fig. 1. Equivalent VRS embedding: allocation cost S=17 (b) and S=8 (c).

II. EMBEDDING VIRTUAL ROUTER SERVICES

Our architecture is based on the assumption that customers expect the functionality of a virtual router service to be indistinguishable from that of a physical device, i.e. the traffic flow between any two nodes attached to the router is limited only by the capacity of their interfaces and routing tables are calculated by a single routing process. Geographical attachment of customer PoPs, corresponding capacity demands as well as the available bandwidth in the substrate are the primary constraints for VRS.

In the following we consider bandwidth allocation costs for a VRS connecting a set of customer PoPs N with capacity demands b_u for $u \in N$.

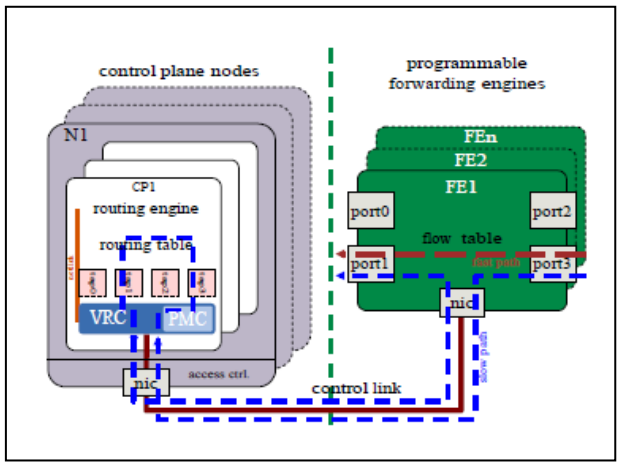


Fig. 2 : Virtual Router Service Architecture.

We define the VRS allocation cost S as the sum of reserved substrate bandwidths b, weighted by the respective link costs c. Without loss of generality, we analyze a fully connected substrate topology spanned between $n = |N|$ edges. In terms of capacity the VRS instances depicted in Figures 1b and 1c offer equivalent connectivity. In Fig.1b, $\min(b_u, b_v)$ units of bandwidth are reserved between each pair of nodes $(u, v) \in N$. Hence, the allocation cost S_{full} is

$$S_{full} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \min(b_i, b_j) c_{ij} \qquad (1)$$

Setting the capacity demands and link costs to one, it is evident that the cost increase is quadratic: $S^1_{full} = n(n-1)/2$. Hence the use of a point to point VRS allocation scheme is problematic even for relatively small numbers of PoPs. On the other hand, if we select any node $k \subset N$ and route traffic from all remaining edges over it, as depicted in Fig.1c, the allocation cost becomes

$$S_{star} = \sum_{i=1}^{n-1} \min(b_i, b_k) c_{ik} \qquad (2)$$

Evidently, S_{star} grows linearly with the number of customer edge nodes and $S_{star} < S_{mesh}$ for all $k \in N$. In fact, for an appropriately chosen core node k, a star topology provides the overall least cost connectivity between a set of edge nodes in an arbitrary substrate network. We address the optimal location of k in section IV.

III. VIRTUAL ROUTER SERVICE ARCHITECTURE

Based on the cost considerations above we propose a star architecture comprised of a single core node, responsible for all Layer 3 routing decisions, connected to a set of customer edge gateways (CEG) over a series of intermediate nodes (IN). Each VRS is associated with a unique control plane instance running in a virtual machine (VM) hosted at a suitable network site. Each VM controls its associated network elements over a dedicated link. The architecture is illustrated in Fig.2 and an exemplary topology is depicted in Fig.3.

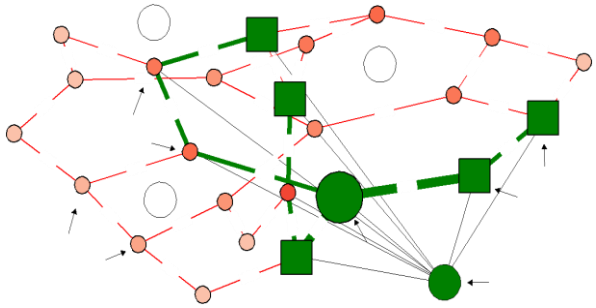


Fig. 3 : Minimum cost VRS allocation (S = 10): customer edge nodes $E = \{3, 7, 11, 14, 22\}$ connected to a VR core node $r = \{5\}$ over intermediate nodes $I = \{1, 10, 6, 3, 11\}$.

The architecture relies on a programmable network substrate which allows a VRS controller to modify the L2 and L3 flow tables of all associated forwarding engines. We believe, that the deployment of programmable network devices will become widespread with the emergence of the Openflow specification [5] which allows external programming of forwarding tables in commercial switches.

The VRS core node is build around the virtual router architecture we proposed in [6]. A virtual router controller (VRC) executed within each VRS control plane transparently installs L3 forwarding rules

needed to mirror the VR routing tables to an Openflow compliant forwarding engine. We extend the architecture to include the setup and management of paths connecting the VR core to customer PoPs at the network edge. The selection of least cost paths and the installation of the corresponding forwarding entries is performed by a path management controller (PMC) (see Fig.2). We exploit the fact that all traffic belonging to a specific customer is identifiable by the L2 addresses of the associated VR. At each IN a forwarding entry directs packets towards the next hop to the core node, based on the L2 destination address of the corresponding VR port. Similarly, for packets exiting the VR core, we use the L2 source address to setup a forwarding path in the reverse direction.

IV. ALGORITHMS FOR VIRTUAL ROUTER SERVICE EMBEDDING

Embedding a VRS involves two independent operations: the selection of an optimal core node location and the allocation of optimal forwarding paths to the CEGs.

Path Selection: The goal of the operation is the identification of least cost paths connecting the core node $\mathbf{r}$ to a set of CEGs E while providing sufficient capacity. If substrate capacity contraints are ignored, reserving bandwidth along the shortest paths from $\mathbf{r}$ to $e \in E$ yields the minimum VRS allocation cost. Fulfilling edge demands in uncapacitated networks is discussed in [7] in the context of point-to-cloud VPNs. However, substrate network capacity is generally limited. In this case, a basic shortest path approach is not guaranteed to minimize the allocation cost.

We formulate the VRS path allocation task as a flow network problem, which can be solved using a minimum cost flow (MCF) algorithm.

Algorithm 1 VRS embedding

1: prune nodes with insufficient resources

2: $S_\infty \leftarrow \infty$ // initialize array of lower bound costs

3: for $e \in E$ do // iterate through all edge nodes

4: get shortest path distances d(n) from e to all $n \in G_\infty$

5: $S_\infty(n) \leftarrow S_\infty(n) + d(n)b(e)$

6: end for

7: sort S_∞ by ascending cost

8: $s_{min} \leftarrow \infty$, $r_{min} \leftarrow \emptyset$

9: $(n, s) \leftarrow pop_0(S_\infty)$ // remove least cost node/cost tuple

10: while $s > s_{min}$ do

11: $s_{min} \leftarrow SSP(n,G)$, $r_{min} \leftarrow n$

12: $(n, s) \leftarrow pop_0(S_\infty)$

13: end while

14: return r_{min}, s_{min}

We interpret the substrate graph G as a flow network and define the CEGs as traffic sinks with a flow demand of b_e and the core node $\mathbf{r}$ as a traffic source with a flow supply of $b_{\mathbf{r}} = -\sum_E b_e$. An optimal set of paths w.r.t. to any given core $\mathbf{r}$ can be calculated using the successive shortest paths (SSP) [8] algorithm, among others. The SSP algorithm has the advantage that it can efficiently handle edge demand changes or attachment of new CEGs. Note that the optimal flow may be split along multiple paths as proposed in [2] if demands $b_i = b_j$ for $(i, j) \in E$.

Core Node Selection: The choice of the core node location is vital to ensure a minimum cost VRS allocation. To avoid checking every feasible core node candidate for optimality us- ing the SSP algorithm, we consider the uncapacitated instance of the substrate graph G_∞. We then calculate the allocation costs $S_\infty(n)$ for all $n \in G_\infty$ using Dijkstra's algorithm and use these as a lower bound for the capacity constrained case as in Alg. 1. Our simulations confirm that this approach substantially reduces the number of required iterations.

Resource constraints: The search space can be significantly reduced by pruning substrate nodes and links with insufficient resources such as bandwidth, switching capacity or forwarding table space. Furthermore, latency bounds between VRS edges can be enforced by eliminating core node candidates exceeding predefined SLAs.

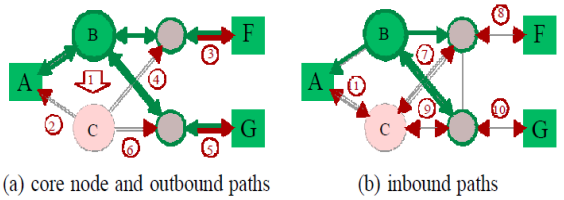
V. DATAPATH LIVE-MIGRATION

A strength of the VRS abstraction is the ability to trans- parently migrate logical entities between physical resources, eliminating the need for reconfiguration. We outline a data plane migration procedure which uses the programmability of the network substrate to migrate VRS across multihop paths. Hence, our approach differs from [9].

First, a new core node $\mathbf{r}^*$ is selected and

corresponding forwarding paths are calculated by the PMC. The VRC then clones and continuously updates L3 forwarding entries at r and r^*. Subsequently, outbound paths are established from r^* towards the edge gateways $e \in E$ of the VRS (Fig.4a).

To ensure connectivity, forwarding entries at each hop are installed beginning from the edge towards the core. Next, starting from the core to the edges, all inbound paths towards r^* are asynchronously established (Fig.4b).



(a) core node and outbound paths (b) inbound paths

Fig. 4. VRS dataplane migration: core node B → C.

As outbound paths are already setup, packet loss cannot occur. However, packets may arrive out of order due to differences between old and new paths lengths. Finally, the original forwarding entries are torn down after all inbound paths have been redirected. The process is fully transparent from a customer's perspective.

VI. CONCLUSION

We outlined an architecture for virtual router services which transparently manipulates the forwarding tables of a set of distributed devices allowing them to be operated as a single entity. The VRS takes advantage of the programmability offered by state-of-the-art network components. By defining the VRS in terms of customer edge capacity demands, the calculation of optimal substrate mappings is made possible. We presented algorithms for a minimum cost VRS embedding in capacity constrained substrate networks. The ability to efficiently allocate VRS instances and migrate resources on the fly paves the way for attractive new business models while ensuring a simplified deployment and operation.We implemented and evaluated a prototype of the presented VRS architecture in our network testbed. Ongoing work includes the distribution of the core node functionality to multiple hosts and the accommodation of additional customer requirements.

REFERENCES

[1] Y. Zhu and M. Ammar, "Algorithms for assigning substrate network resources to virtual network components," in Proc. IEEE INFOCOM 2006, Apr. 2006, pp. 1 –12.

[2] M. Yu, Y. Yi, J. Rexford, and M. Chiang, "Rethinking virtual network em- bedding: substrate support for path splitting and migration," SIGCOMM CCR, vol. 38, pp. 17–29, March 2008.

[3] N. Chowdhury, M. Rahman, and R. Boutaba, "Virtual network embedding with coordinated node and link mapping," in Proc. IEEE INFOCOM 2009, Apr. 2009, pp. 783 –791.

[4] E. Keller and J. Rexford, "The "platform as a service" model for networking," in Proc. INM/WREN '10, April 2010.

[5] N. McKeown et. al., "Openflow: enabling innovation in campus net- works," SIGCOMM CCR, vol. 38, no. 2, pp. 69–74, 2008.

[6] Z. Bozakov, "An open router virtualization framework using a pro- grammable forwarding plane," in Proc. ACM SIGCOMM, 2010, pp. 439–440.

[7] N. Duffield et. al., "Resource management with hoses: point-to-cloud ser- vices for virtual private networks," Networking, IEEE/ACM Transactions on, vol. 10, no. 5, pp. 679 – 692, Oct. 2002.

[8] R. K. Ahuja, T. L. Magnati, and J. B. Orlin, Network Flows: Theory, Algorithms, and Applications. Prentice Hall, 1993, pp. 320–324.

[9] Y. Wang et. al., "Virtual routers on the move: live router migration as a network-management primitive," in Proc. ACM SIGCOMM, 2008, pp.231–242.

◈◈◈

Speaker Recognition

Using K-Means Algorithm

Kavita Bhangale & Shoba Krishnan

Dept. of Electronics & Tele Communication, Vivekanand Education Society Institute of Technology, Mumbai, India
E-mail : kavita_bhangale@rediffmail.com, Shobakrishnan87@hotmail.com

Abstract - This paper presents recognition of speaker from a variety of database given. . Mel frequency Cepstral Coefficients{MFCCs} have been used for feature extraction and vector quantization technique is used to minimize the amount of data to be handled.

Key words - *speaker recognition, MFCC,K-means algorithm.*

I. INTRODUCTION

Speech is one of the natural forms of communication. Recent development has made it possible to use this in the security system. In speaker identification, the task is to use a speech sample to select the identity of the person that produced the speech from among a population of speakers. In speaker verification, the task is to use a speech sample to test whether a person who claims to have produced the speech. This technique makes it possible to use the speakers' voice to verify their identity and control access to services such as voice dialing, banking by telephone, telephone shopping, database access services, information services, voice mail, security control for confidential information areas, and remote access to computers.

II. PRINCIPLES OF SPEAKER RECOGNITION

Speaker recognition methods can be divided into text-independent and text-dependent methods. In a text-independent system, speaker models capture characteristics of somebody's speech which show up irrespective of what one is saying. In a text-dependent system, on the other hand, the recognition of the speaker's identity is based on his or her speaking one or more specific phrases, like passwords, card numbers, PIN codes, etc. Every technology of speaker recognition, identification and verification, whether text-independent and text dependent, each has its own advantages and disadvantages and may require different treatments and techniques. The choice of which technology to use is application-specific. At the highest level, all speaker recognition systems contain two main modules feature extraction and feature matching.

III. SPEECH FEATURE EXTRACTION

The purpose of this module is to convert the speech waveform to some type of parametric representation (at a considerably lower information rate). The speech signal is a slowly time varying signal (it is called quasi-stationary). When examined over a sufficiently short period of time (between 5 and 100 ms), its characteristics are fairly stationary. However, over long periods of time (on the order of 0.2s or more) the signal characteristics change to reflect the different speech sounds being spoken. Therefore, short-time spectral analysis is the most common way to characterize the speech signal. A wide range of possibilities exist for parametrically representing the speech signal for the speaker recognition task, such as Linear Prediction Coding (LPC), Mel-Frequency Cepstrum Coefficients (MFCC), and others. MFCC is perhaps the best known and most popular, and this feature has been used in this paper. MFCC's are based on the known variation of the human ear's critical bandwidths with frequency. The MFCC technique makes use of two types of filter, namely, linearly spaced filters and logarithmically spaced filters. To capture the phonetically important characteristics of speech, signal is expressed in the Mel frequency scale. This scale has a linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz. Normal speech waveform may vary from time to time depending on the physical condition of speakers' vocal cord. Rather than the speech waveforms themselves, MFFCs are less susceptible to the said variations .For the feature extraction section, the used algorithm is calculating the Mel-Frequency Cepstral Coefficients (MFCC).The aim of this feature extraction process is to obtain a new voice representation which is

more compact, less redundant, and more suitable for statistical modeling. The MFCC is based on the known variation of the human ear's critical bandwidths with frequency, where it filters the space linearly at low frequencies and logarithmically at high frequencies. It is used in order to capture the phonetically imimportant characteristics of the voice. There are several steps in order to implement the MFCC as shown in the Figure 1.

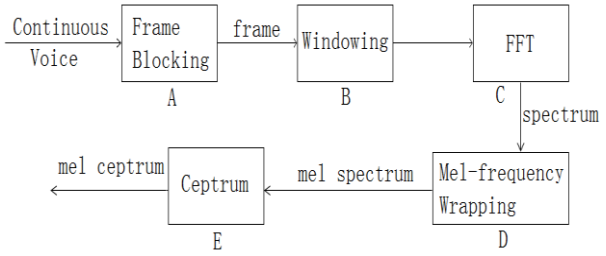


Fig. 1 : Computation of Mel Frequency Cepstral Coefficients (MFCC).

Process A: Frame Blocking

The framing process is firstly applied to the voice signal of the producer. This signal is partitioned or blocked into N segments .(frames).

Process B: Windowing

The second process of the processing is to window each of the individual frame such to minimize the signal discontinuities at the beginning and end of each frame.

Process C: Fast Fourier Transform

The next process is the Fast Fourier Transform (FFT) where each frame of *N* samples is converted from time domain to frequency domain.

Process D: Mel-Frequency Wrapping

The obtained spectrum from the FFT process is then Mel Frequency Wrapped. The major aim of this process is to convert the frequency spectrum to the Mel spectrum.

Process E: Cepstrum

In the final process, the log Mel spectrum is then converted back to time domain and the result is called the Mel frequency Cepstrum Coefficients (MFCC).

Vector quantization

Vector quantization (VQ) is a lossy data compression method based on principle of block coding. It is a fixed-to-fixed length algorithm. VQ may be thought as an aproximator. Figure 2 shows an example of a 2-dimensional VQ. Here, every pair of numbers falling in a particular region are approximated by a star associated with that region. In Figure 2, the stars are called codevectors and the regions defined by the borders are called encoding regions.

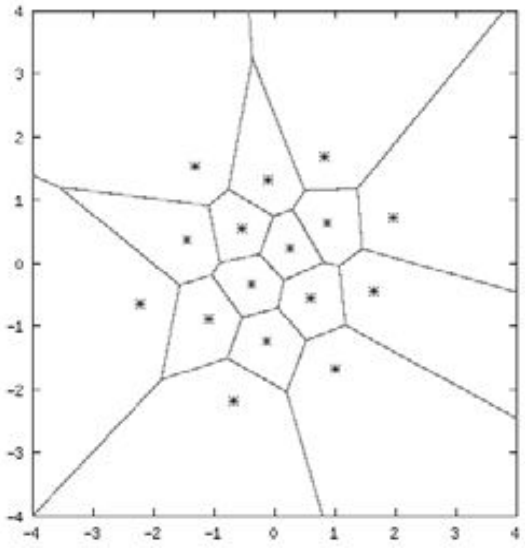


Fig.2 **:** An example of a 2-dimensional VQ

The set of all codevectors is called the codebook and the set of all encoding regions is called the partition of the space .In the training phase, a speaker-specific VQ codebook is generated for each known speaker by clustering his/her training acoustic vectors. The resultant codewords (centroids) are shown in Figure 3 by circles and triangles at the centers of the corresponding blocks for speaker1 and 2, respectively. The distance from a vector to the closest codeword of a codebook is called a VQ distortion. In the recognition phase, an input utterance of an unknown voice is "vector-quantized" using each trained codebook and the *total VQ distortion* is computed. The speaker corresponding to the VQ codebook with the smallest total distortion is identified.

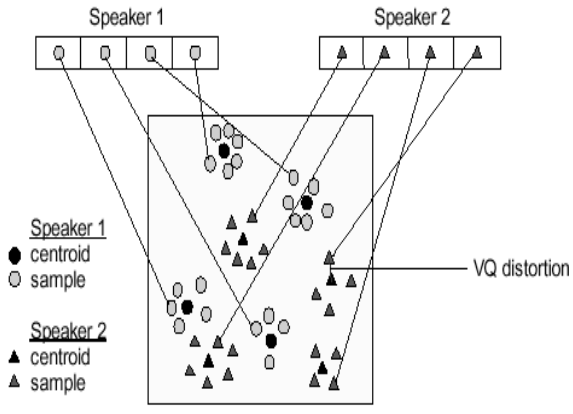


Fig. 3 **:** Conceptual diagram to illustrate the VQ Process.
K-Means algorithm

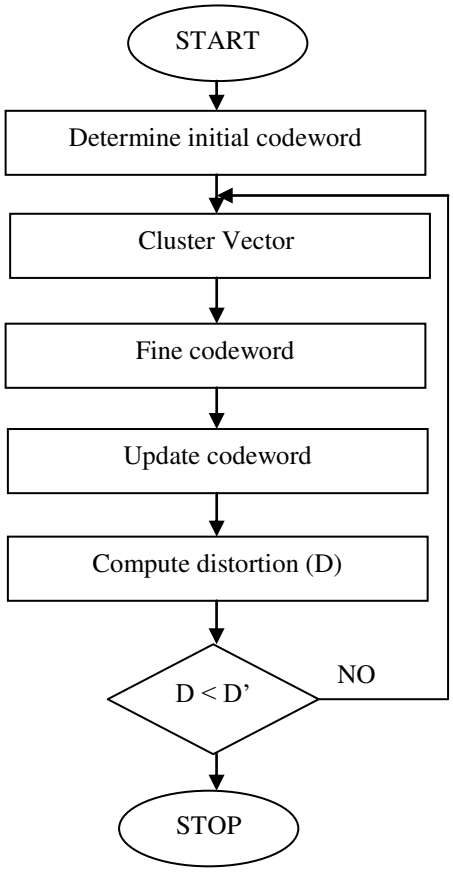


Fig.4 : k-means algorithm

k-means algorithm

This part briefly describes the standard k-means algorithm. Kmeans is a typical clustering algorithm in data mining and which is widely used for clustering large set of datas. In 1967, MacQueen firstly proposed the k-means algorithm, it was one of the most simple, non-supervised learning algorithms, which was applied to solve the problem of the well-known cluster . It is a partitioning clustering algorithm, this method is to classify the given date objects into k different clusters through the iterative, converging to a local minimum. So the results of generated clusters are compact and independent. The algorithm consists of two separate phases. The first phase selects k centers randomly, where the value k is fixed in advance. The next phase is to take each data object to the nearest center. Euclidean distance is generally considered to determine the distance between each data object and the cluster centers. When all the data objcets are included in some clusters, the first step is completed and an early grouping is done. Recalculating the average of the early formed clusters. This iterative process continues repeatedly until the criterion function becomes the minimum. Supposing that the target object is x, xi indicates the average of cluster $C_{i,}$ criterion function is defined as follows:

$$E = \sum_{i=1}^{k} \sum_{x \in Ci} \left| x - x_i \right|^2$$

E is the sum of the squared error of all objects in database. The distance of criterion function is Euclidean distance, which is used for determining the nearest distance between each data object and cluster center. The Euclidean distance between one vector x=(x1 ,x2 ,…xn) and another vector y=(y_1 ,y_2 ,…y_n), The Euclidean distance d(x_i, y_i) can be obtained as follow:

$$d(x_i, y_i) = \{ \sum_{i=1}^{n} (x_i - y_i)^2 \}^{1/2}$$

IV. SPEAKER MATCHING

In the recognition phase an unknown speaker, represented by a sequence of feature vectors is compared with the codebooks in the database. For each codebook a distortion measure is computed, and the speaker with the lowest distortion is chosen. One way to define the distortion measure is to use the average of the Euclidean Distances. The Euclidean distance is the ordinary distance between the two points that one would measure with a ruler, which can be proven by repeated application of the Pythagorean Theorem. Thus, each feature vector in the sequence X is compared with all the codebooks, and the codebook with the minimized average distance is chosen to be the best.

V. RESULT:

Loading data...

Calculating mel-frequency cepstral coefficients for training set...

a1

b1

c1

d1

Performing K-means...

Calculating mel-frequency cepstral coefficients for test set...

Compute a distortion measure for each codebook...

Display the result...

The average of Euclidean distances between database and test wave file

a1

6.0992

b1

16.0344

c1

12.0239

d1

6.7690

The test voice is most likely from

a1

VI. CONCLUSION

Using speaker recognition system , we can identify the person who is speaking. In the recognition stage, a distortion measure which based on the minimizing the Euclidean distance was used when matching an unknown speaker with the speaker database.

REFERENCES

[1] Lawrence Rabiner and Biing-Hwang Juang, Fundamental of Speech Recognition", Prentice-Hall, Englewood Cliffs, N.J., 1993.

[2] Zhong-Xuan, Yuan & Bo-Ling, Xu & Chong-Zhi, Yu. (1999). "Binary Quantization of Feature Vectors for Robust Text-Independent Speaker Identification" in IEEE Transactions on Speech and Audio Processing, Vol. 7, No. 1, January 1999. IEEE, New York, NY, U.S.A.

[3] F. Soong, E. Rosenberg, B. Juang, and L. Rabiner, "A Vector Quantization Approach to SpeakerRecognition", AT&T Technical Journal, vol. 66,March/April 1987, pp. 14-26

[4] Comp.speech Frequently Asked Questions WWW site,

http://svr-www.eng.cam.ac.uk/comp.speech/

[5] R. M. Gray, ``Vector Quantization," IEEE ASSP Magazine, pp. 4--29, April 1984.

[6] Y. Linde, A. Buzo & R. Gray, "An algorithm for vector quantizer design", IEEE Transactions on Communications, Vol. 28, pp.84-95, 1980.

[7] Jr., J. D., Hansen, J., and Proakis, J. Discrete-Time Processing of Speech Signals, second ed. IEEE Press, New York, 2000.

[8] The Comparison of Vector Quantization Algoritms in Fish Species Acoustic Voice Recognition Using Hidden Markov Model Diponegoro A.D1). and Fawwaz Al Maki

◈◈◈

```
%function speakerID(test)
% A speaker recognition program. a is a string
of the filename to be tested
% against the database of sampled voices and it
will be evaluated whose
% voice it is.
% - Example -
% to test a 'test.wav' file then,
% >> speakerID('test.wav')
%
% - Reference -
% Lasse Molgaard and Kasper Jorgensen, Speaker
Recognition,
%
www2.imm.dtu.dk/pubdb/views/edoc_download.php/4
414/pdf/imm4414.pdf
%
% Mike Brooks, VOICEBOX, Free toolbox for
MATLab,
% www.ncl.ac.uk/CPACTsoftware/MatlabLinks.html
% disteusq.m enframe.m kmeans.m melbankm.m
melcepst.m rdct.m rfft.m from
% VOICEBOX are used in this program.

clear all;clc;
test.data = wavread('test');
% read the test file
name = ['a1';'b1';'c1';'d1'];   % name of
people in the database

fs = 16000;              % sampling frequency
C = 8;                   % number of centroids

% Load data
disp('Loading data...')
[train.data] = Load_data(name);

% Calculate mel-frequecny cepstral coefficients
for training set
disp('Calculating mel-frequency cepstral
coefficients for training set...')
[train.cc] = mfcc(train.data,name,fs);

% Perform K-means algorithm for clustering
(Vector Quantization)
disp('Performing K-means...')
[train.kmeans] = kmean(train.cc,C); % Calculate
mel-frequecny cepstral coefficients for
training set
disp('Calculating mel-frequency cepstral
coefficients for test set...')
test.cc = melcepst(test.data,fs,'x');

% Compute average distances between test.cc
with all the codebooks in
% database, and find the lowest distortion
disp('Compute a distortion measure for each
codebook...')
[result index] =
distmeasure(train.kmeans,test.cc);

% Display results - average distances between
the features of unknown voice
% (test.cc) with all the codebooks in database
and identify the person with
% the lowest distance
disp('Display the result...')
dispresult(name,result,index)
```

Implementation of Modified Architecture for Adaptive Viterbi Decoder

Pallavi R. More & Divya M. Meshram

Department of Electronics Engineering, PCE, Nagpur, India
E-mail : pallavimorey@gmail.com, divyameshram@gmail.com

Abstract **-** The demand for high speed, low power and low cost for Viterbi decoding especially in wireless communication are always required. Thus the paper presents the design of an adaptive Viterbi decoder that uses survivor path with parameters for wireless communication in an attempt to reduce the power and cost and at the same time increase the speed. Nowadays, most digital communication systems convolutionally encoded the transmitted data to compensate for Additive White Gaussian Noise (AWGN), fading of the channel, quantization distortions and other data degradation effects. For its efficiency the Adaptive viterbi algorithm has proven to be a very practical algorithm for forward error correction of convolutionally encoded messages. The requirements for the Adaptive Viterbi decoder depend on the applications used. It plays an important role in the future communication system such as 3rd Generation Partnership Project and DVB. The Adaptive Viterbi Decoder uses adaptive Viterbi algorithm to reduce the average computation and path storage required by the Viterbi algorithm. When Adaptive viterbi decoder operates in the real system, It affects the system performance By increasing speed of operation, reducing power consumption and cost.

I. INTRODUCTION

As mobile and wireless communication becomes increasingly ubiquitous, the need for dynamic reconfigure ability of hardware shall pose fundamental challenges for communication algorithm designers as well as hardware architectures. A new Adaptive Viterbi decoder architecture has been proposed. It is generally used in wireless communication system for high speed, low power and low cost. Instead of computing and retaining all $2K-1$ possible paths, only those paths which satisfy certain cost conditions are retained for each received symbol at each state node. The architecture is simple and suitable for very high date rate decoding, reconfigurable the Viterbi decoders units adding a unit called Viterbi ROM. The structure is used to store 21 new storage unit used to store the correct paths.

II. CONVOLUTIONAL ENCODER

The convolutional Encoder provides powerful error correcting & encoding capability. It offers an alternative to block codes for transmission over a noisy channel Encoded bits are functions of information bits and the number of memory elements. The information sequence is shifted into and along a shift registers *k* bits at a time. Bits are tapped off at different stages of the shift register and summed in a modulo-2 adder (XOR gate).

1. The convolutional encoder is defined by two parameters: Constraint Length K = Number of shift register + 1. This basically represents the number of Locations from where bits can be tapped of.
2. Rate = k/n. If there are n modulo 2 adders, it means that for every k-bit shift, there will be an output of k-bits.

Most wireless applications define the rate to be 1/2., which means that for every one bit that enters the convolutional encoder, 2 bits are received at the output. For the proposed design, the constraint length is kept at 3; therefore there will be two shift registers at the encoder side.

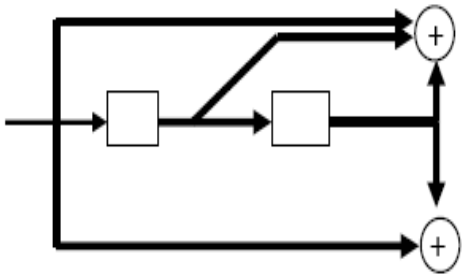


Fig. 1 : 4 States Convolutional Encoder

III. STATE DIAGRAM

A convolutional encoder is a Mealy machine, where the output is a function of the current state and the current input. It consists of one or more shift registers and multiple XOR gates. The operation of a convolutional encoder can be easily understood with the aid of a state diagram. Figure 2 represents the state diagram of the encoder shown in Figure 2. It also

depicts state transitions and the corresponding encoded outputs.

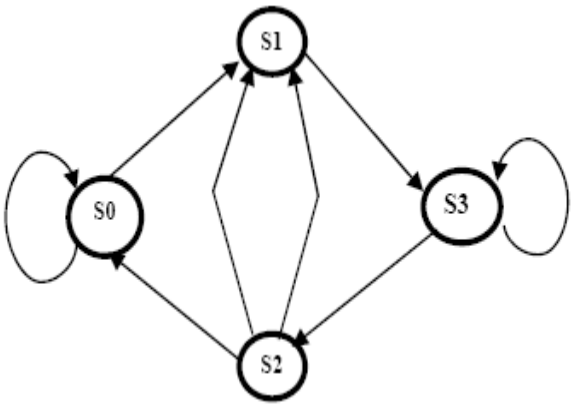


Fig. 2 : State Diagram for Convolutional Encoder

As there are two memory-elements in the circuit, there are four possible states that the circuit can assume. These four states are represented as S0 through S3. Each state's information (i.e. the contents of flip-flops for the state) along with an input generates an encoded output code. For each state, there can be two outgoing transitions; one corresponding to a '0' input bit and the other corresponding to a '1'input bit.

IV. THE TRELLIS DIAGRAM

A trellis diagram is an extension of a state diagram that explicitly shows the passage of time. Figure 3 shows a trellis diagram for the encoder.

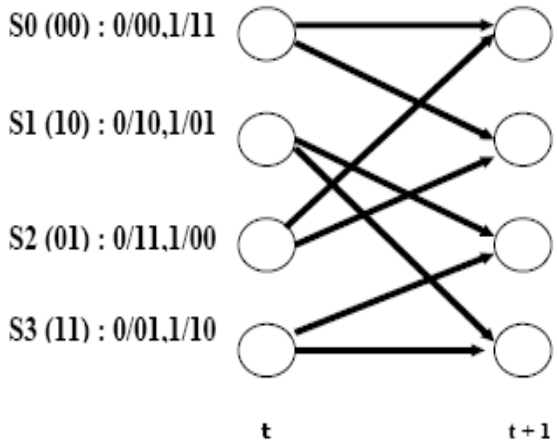


Fig. 3 : Trellis Diagram for the Convolutional Encoder

In the trellis diagram, nodes correspond to the states of the encoder. From an initial state (S0) the trellis records the possible transitions to the next states for each possible input pattern. For the encoder in Figure 2, there are two encoded symbols corresponding to input bit '0' and '1'. The Figure 3 shows the encoded symbol generated for each transition. A the stage t=1 there are two states S0 and S1, and each state has two transitions corresponding to input bits '0' and '1'. Hence the trellis grows up to the maximum number of states or nodes, which is decided by the number of memory elements in the encoder.

V. ADAPTIVE VITERBI DECODER (AVD)

The aim of the adaptive Viterbi algorithm is to reduce the average computation and path storage required by the Viterbi algorithm. Instead of computing and retaining all $2K-1$ possible paths, only those paths which satisfy certain cost conditions are retained for each received symbol at each state node. Path retention is based on the following criteria.

1. A threshold T indicates that a path is retained if its path cost is less than dm + T, where dm is the minimum cost among all surviving paths in the previous trellis stage.
2. The total number of survivor paths per trellis stage is limited to a fixed number, Nmax, which is pre set prior to the start of communication.

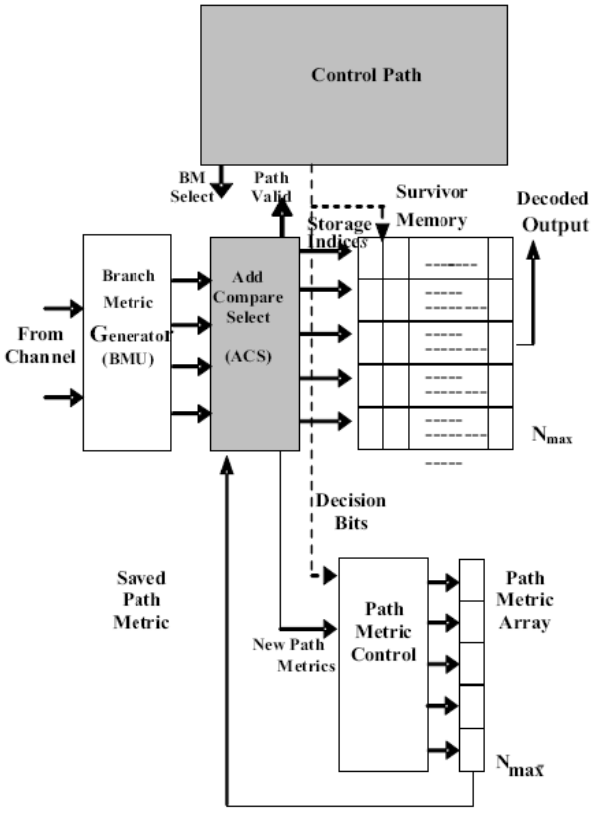


Fig. 4 : Adaptive Viterbi Decoder Architecture

A high level view of adaptive Viterbi decoder architecture is shown in Fig. 1. The decoder contains a data path and an associated control path. Like most Viterbi decoder, the data path is split into four parts: the Branch Metric Generators (BMG), Add Compare Select (ACS) units, the survivor memory unit, and path metric storage and control. A BMG unit determines path costs and identifies lowest cost paths. The survivor memory stores lowest cost bit sequence paths on decisions made by the ACS units, and the path metric array holds per state path metrics. The flow of data in the data path and the storage of results are determined by the control path.

VI. SURVIVOR PATH STORAGE BLOCK

This block is necessary only for the trace back algorithm. However, the algorithm considers the forward paths to find the survivor path. The following steps show the operations used to find the survivor path:

Step 1: Let *P* be the length of the path. *P* is the number of the symbols read in the received side at each Viterbi computation.

Step 2: The size of the RAM depends on the length of the Path.

Step 3: The index of the RAM represents the code of the survivor path but in the opposite direction. If the index Value is (10)10, the code of this value in binary (1010) 2 and the survivor path is (0101) which represents the output code.

Step 4: The first, second,... *P*th columns of the RAM store the values of the Hamming distance of each path calculated.

Step 5: The next column is used to add the Hamming of the path stored in the columns explained in b of each row.

Step 6: The last column is used to store the number of the next state that the path may take in the trills diagram for this code.

Step 7: The size of the RAM is $2P*(P+2)$. The first column is used to store the Hamming distance of each selected path which can be represented by 2 bits only. This is because the maximum Hamming distance between the expected code and the receive symbols is 3 that can be represented by 2 bits only, while the next columns need only 4 bits to store the total maximum Hamming distance of 4 columns. This will give the maximum number stores in this column, 3*4=12. This can be represented by 4 bits only. The last column stores the number of the next state in the trills diagram. For a decoder of (3, 1, 7), the number of states in the trills diagram are 64 states (27-1 states). Therefore, only 6 bits are required to represent this number and the total bits required of each rows in the RAM are 16 bits.

Step 8: The length of the trills is equal to the length of the input sequence. This consists of the input information bits followed by the reset sequence. The number of the reset sequence is the same as the number of memory elements used in the encoder. This reset sequence gives an indication to the designer that, it is the end of the symbol and the symbol start at the state *S0* and ending at the same state *S0*. Thus the next Symbol started again with the same state *S0* and repeated the algorithm to find the next code from the trills diagram. Two approaches are often used to record survivor path, serial and parallel operations. In the serial approach, there is only one RAM to record the computation data of all the paths depending on the path length and it uses the same hardware to find the data.

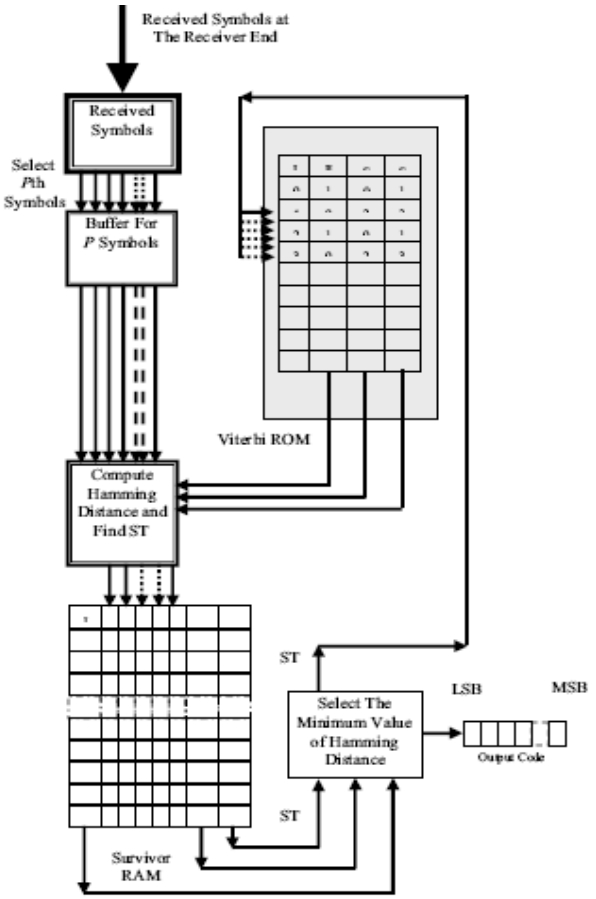


Fig. 5 : Block diagram of survivor path serial algorithm

VII. VITERBI DECODER WITH SURVIVOR PATH STORAGE

The Viterbi decoder with 64 states, 1/3 rate, and a polynomial of (1178, 1278, and 1558) has been used.Using six memory units, the polynomials equations of this encoder can be represented by the following equations:

$$G_0 = ip \otimes D_3 \otimes D_2 \otimes D_1 \otimes D_0 \quad (1)$$

$$G_1 = ip \otimes D_3 \otimes D_2 \otimes D_1 \otimes D_0 \quad (2)$$

$$G_2 = ip \otimes D_3 \otimes D_2 \otimes D_1 \otimes D_0 \quad (3)$$

The block diagram of the internals of the reconfigurable Viterbi decoder consist of ROM's, branch metric unit (BMU), compare select unit (CSU), state metric unit (SMU), and RAM.

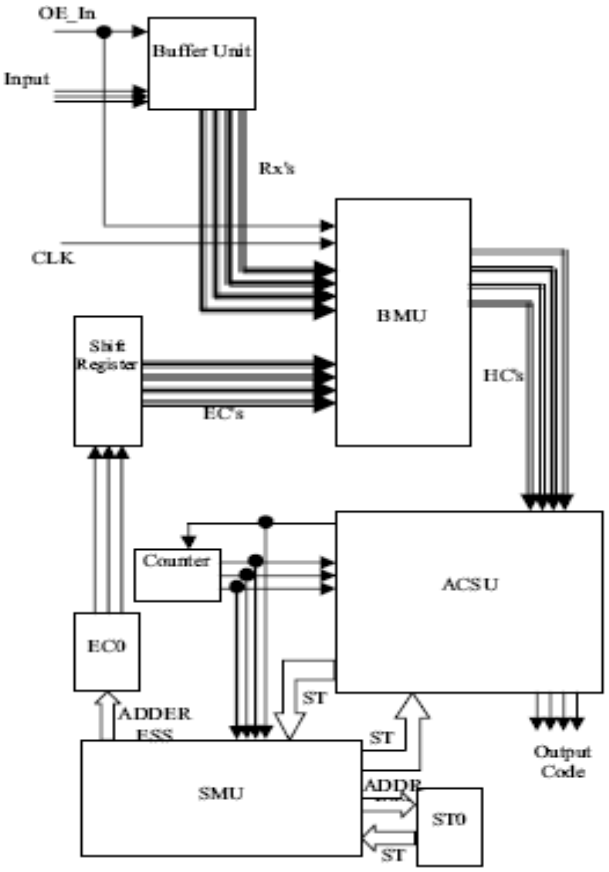


Fig. 6 : Block Diagram of Survivor Viterbi Decoder

1. Input / Output Data: The input data of the decoder are three parallel bits represents the information data and the control signal of the Viterbi decoder processing with the CLK which is a standard input signal to the input port of the decoder. Furthermore, the OE_In is a control signal that enables the input of the decoder and the last input is the OE_out control signal that enables the output of the decoder.
2. Architecture of Buffer Unit: The survivor algorithm needs to split the received data frame into sub fame. Each sub frame has (*P*) symbols. The (*P*) symbols are stored in the buffers for comparison with all probability paths that take the same the RAM index.
3. The Survivor Path Storage Memory (RAM): Here, the RAM is splited in to 6 banks. The first four banks are used to store the Hamming code resulted from the comparison between the codes stored in the buffer with the path represents the address RAM. The fifth bank stores the addition of the total Hamming code of the *P* symbols which represent the Hamming code of each raw. The last bank is used to store the next state number.
4. The Branch Metric Unit (BMU): The responsibility of this unit is to compute the Hamming code between the expected code and the receiving code as a frame. Each frame contains four symbols. At each processing, the BMU finds the Hamming code for these four symbols. This will be compared with the expected code represented by the address value that replaced by a counter started with '0000' to '1111'. The followings explain this operation:

Step 1: In the first frame and starting with ST=0; the BMU Computes Hamming code between, Rx0 and EC0 (ST). The variable between these two arches represent the value of the address ROM.

Step 2: STnew = ST0 (STold).

Step 3: Next the Hamming code between Rx1 and the value stored in the EC0 (ST) is computed.

Step 4: STnew = ST0 (STold).

Step 5: The Hamming code between Rx2 and the value stored in the EC0 (ST) is computed.

Step 6: ST0 (STold).

Step 7: The Hamming code between Rx3 and the value stored in the EC0 (ST) is computed.

Step 8: STnew = ST0 (STold).

5. The Add Compare Select Unit (ACSU): This ACSU is the main unit of the survivor path decoder. The function of this unit is to find the addition of the four Hamming code received from BMU's and to compare the total Hamming code stored in the HC_buffer for the even and odd paths. This unit also compares the values and store, the minimum Hamming code, and the counter's value. At the end of the 8th even state, the minimum Hamming code and the counter value are stored in HC_buffer and AD_buffer. The AD_buffer represents the opposite output code. The other value stored in ST_buffer represents the address of the next state number.

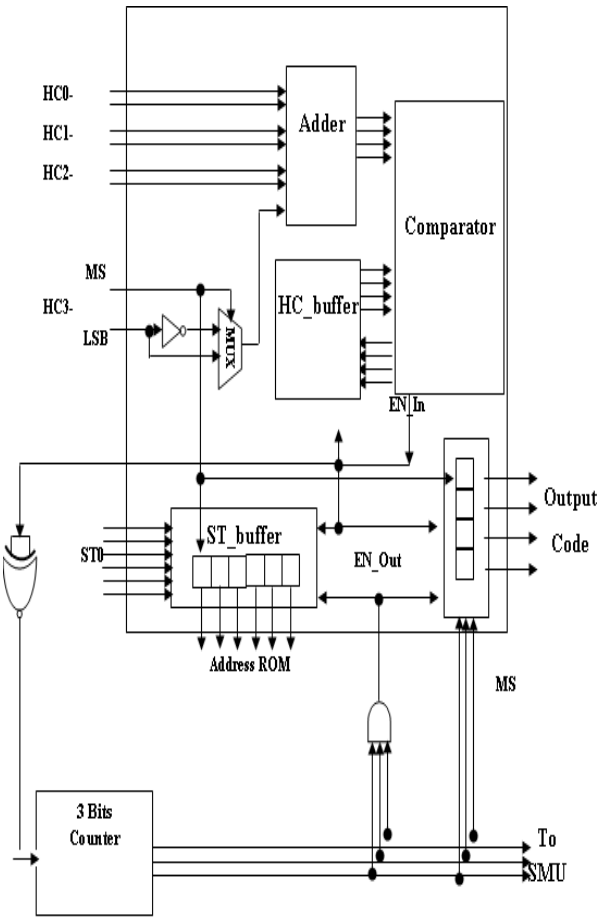


Fig. 7 : Block diagram of ACSU

VIII. ADVANTAGES

- The use of error-correcting codes has proven to be an effective way to overcome data corruption in digital communication channels.
- High speed
- Low power
- Low cost

IX. APPLICATIONS OF VITERBI DECODER

- A Low-Power Viterbi Decoder for Wireless Communications Applications.
- Pipelined VLSI Architecture of The Viterbi Decoder for IMT-2000.
- 200Mbps Viterbi decoder for UWB.
- Viterbi Decoder for WCDMA System.
- Low complexity efficient trackback viterbi decoder for wireless applications.
- A Soft IP Generator for High-speed Viterbi Decoders.
- Suitable Mobile Channel Conditions for a Concatenated Coding System with List-of-2 Viterbi Inner Decoder.
- High-Speed Low-Power Viterbi Decoder Design for TCM Decoders.
- Low Power Viterbi Decoder by Modified ACSU architecture and Clock Gating Method.

X. CONCLUSION

- As mobile and wireless communication becomes increasingly ubiquitous, the need for dynamic reconfigure ability of hardware shall pose fundamental challenges for communication algorithm designers as well as hardware architectures. This paper attempts to solve this problem for the Particular case of the Viterbi decoder.
- Reconfigure the Viterbi decoder, and adaptive Viterbi Decoder units will give simple elements in each unit and new algorithms.
- The processing execution time has been reduced by removing the trace back algorithms that is used to find the correct paths.
- The survivor path algorithm used, the address of the Memory unit to select the correct path which specifies the output code.

SOFTWARE AND HARDWARE REQUIREMENT

For Software simulation I will prefer MODELSIM and for synthesis I will be prefer XILINX. Hardware requirement is SPARTAN-3.

RESULT VERIFICATION AND ANALYSIS

Observe the required result like adaptive viterbi decoding for different codes.

ACKNOWLEDGEMENTS

Authors wish to remark the great task carried out by the Xilinx and Modelsim user guide; and the authors wish to thank Divya Meshram for his contribution in the design progress.

REFERENCE

[1] "Prototype of a high performanc generic Viterbi decoder" -Obeid A. M., Ortiz A. G., Ludewig R., and Glenser M.

[2] "Design and implementation of low-power Viterbi decoder for software-defined WiMAX receiver" -S. W. Shaker, S. H. Alramely and K. A. Shehata

[3] "FPGA implementation of Viterbi decoder"- H.S, Suresh and B.V, Ramesh.

[4] "A parallel Viterbi decoder for block cyclic and convolutional codes" -J. S, Reeve and K. Amarasinghe

[5] "Implementation of Viterbi encoder and decoder on FPGA" - M. Pramod and M. K. Patil,

[6] VHDL – Cookbook - Peter J. Ashenden

[7] "Asynchronous Viterbi decoder in action Systems", J. Tuominen and J. Plosila., 23rd NORCHIP Conference, 2005.

◈◈◈

Non Linear Controller for DC-DC Buck Converter

T. Chinthu & Poorani Shivkumar

Department of EEE, Sona College of Technology, Salem,Tamil Nadu, India
E-mail : mail2chin@gmail.com, drpooranieee@gmail.com

Abstract - A DC-DC buck converter with constant load is modeled from the state space equations and its dynamic performance is analyzed with a Proportional Integral Derivative, Sliding Mode Controller and Sliding Mode Proportional Integral Derivative Controller respectively. On Comparative analysis of the responses obtained, the Sliding Mode Proportional Integral Derivative controller provides a faster response, thus stabilizing the system efficiently. Sliding Mode Proportional Integral Derivative Controlled converter proves to be efficient and economical in applications requiring power saving criteria's.

Key words - *PID Controller, Sliding Mode Controller, SMCPID Controller, State Space Equation.*

I. INTRODUCTION

A step down type switching regulator, Buck converter finds its applications in Battery powered devices like Cell Phones, Laptop, Electric vehicle, etc, Recently it is also being used in renewable energy processing , as maximum output power can be obtained at higher efficiency. In order to optimize the efficiency and for proper power management, the issues like effect of load variation, power ON transients, Switching and Electromagnetic interference loss has to be overcome for which controllers are used.

In the proposed method, Proportional Integral Derivative, Sliding Mode and Sliding Mode Proportional Integral Derivative controllers are designed independently for a buck converter and the response for appropriate control parameters has been obtained. The stability of the system has been analyzed from the performance characteristics tabulated, which clearly shows that the Sliding Mode Proportional Integral Derivative controlled converter is dynamic and efficient for various applications.

The organization of paper is as follows: Section2 describes the mathematical and state space modeling of Buck converter which is followed by the design of controllers for the Buck Converter in Section 3. Section 4 comprises of the results and comparison table.

II. MODELING OF BUCK CONVERTER

A. Mathematical Modeling

The buck converters operates either in continuous or discontinuous mode depending on the circuit components which includes 2 switches, (i.e. a diode and a transistor) a inductor and a capacitor along with the load for which smooth acceleration control, high efficiency and faster dynamic response is required.

Therefore, the dynamic equation for the converter designed with the R Load for the circuit topology in [1] is:

For continuous conduction mode:

The output of the circuit for ON period is:

$$\frac{d i_L}{dt} = \frac{-R V_{in}}{L} - \frac{V_C}{L} \quad (1)$$

$$\frac{dV_c}{dt} = \frac{i_L}{C} - \frac{V_c}{R} \quad (2)$$

The output of the circuit for OFF period is:

$$\frac{di_L}{dt} = -\frac{V_c}{L} \quad (3)$$

$$\frac{dV_c}{dt} = \frac{i_L}{C} - \frac{V_c}{RC} \quad (4)$$

For Discontinuous Conduction mode:

$$\frac{di_L}{dt} = 0 \quad (5)$$

$$\frac{dV_C}{dt} = \frac{-V_C}{RC} \quad (6)$$

With reference to these equations, the state space averaging technique is performed for better analysis.

B. State Space Modeling

The state space averaging is an approximation technique, which helps in continuous time signal frequency analysis apart from the switching frequency analysis for higher switching frequencies. Though the

original system is linear, the resulting system will be non-linear therefore, this averaging technique gives an ease in representation of transfer functions [1].

Thus from the dynamic equations, the state space equation for the entire switching cycle T is given below:

For Continuous mode:

$$X = AX + BV_{in} \quad (7)$$

$$V_0 = Cx \quad (8)$$

Where,

$$X = \begin{bmatrix} i_L \\ V_C \end{bmatrix} \qquad A = \begin{bmatrix} 0 & -\frac{1}{L} \\ \frac{1}{C} & -\frac{1}{R_C} \end{bmatrix}$$

$$C = (0\ 1) \qquad B = \begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix}$$

For Discontinuous mode:

$$A = \begin{bmatrix} 0 & 0 \\ 0 & -\frac{1}{RC} \end{bmatrix} \qquad C = (0\ 0)$$

using these equations, the transfer function of the system is determined to be:

$$\frac{V_2}{V_1} = \frac{1}{1+s\frac{L}{R}+S\hat{}2LC} \quad (9)$$

III. CONTROLLERS FOR BUCK CONVERTER

In order to overcome the switching losses due to unregulated power supply and core losses affecting the high frequency operation, controllers are widely used so that, regulated output voltage with maximum power can be obtained. Hence, linear and non-linear controllers are designed independently and the converter stability has been analyzed

A. Conventional Proportional Integral Derivative

The most widely used linear industrial controller, Proportional integral derivative is often combined with logic, sequential functions and simple functional blocks to build the automation systems used for energy production, transportation, and manufacturing. It involves three different parameters P, I and D each performing its functions to stabilize the system [2]. The gain parameters are tuned in such a way that the controller reduces the maximum overshoot in the system thereby stabilizing the system effectively based on the error value.

Fig. 1: Represents the Block diagram for PID controlled Buck Converter.

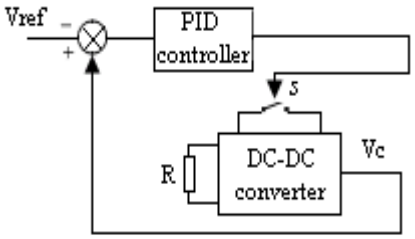


Fig. 1 : Block of PID controlled Buck Converter

B. Sliding Mode Controller

Sliding Mode Controller, a widely used Non-Linear Controller remains vibrant in overcoming the issues of the linear controllers. Sliding mode controller is a variable structure system which operates based on the switching strategy apart from the feedback controllers[3]. They are less sensitive to disturbance and parameter variations due to its binary nature adapting to the modern power switches[4]. The basic principle behind the SMC controlled system is to drive the converter to the steady surface called the sliding surface and maintain the stability of the system thus giving the regulated output voltage for any variations in the load or switching frequency. This depends on the sliding coefficients which determines the trajectory surface. Fig:2 represents the SMC block controlling the buck converter.

The sliding surface is given by:

$$S = Kx_1 + x_2 \quad (10)$$

$$x_2 = V_{ref} - V_0 \quad (11)$$

where, K is the sliding coefficient.

From the modeling equation of the buck converter, the sliding surface is given by:

$$S = \frac{-1}{C} i_L + \left(\frac{1}{CR_L} - K\right) V_C + KV_{ref} \quad (12)$$

Derivative of switching function is given by:

$$\dot{S} = \frac{1}{C}\left(\frac{1}{CR_L}\text{-}K\right) i_L + \left(\frac{\text{-}1}{CR_L}\left(\frac{1}{CR_L}\text{-}K\right) + \frac{1}{LC}\right) V_C - \frac{V_i}{LC} u \quad (13)$$

The SMC control signal u is given by

$$u = \begin{cases} u + (x,t) \text{ for } s > 0 \\ u - (x,t) \text{ for } s < 0 \end{cases} \quad (14)$$

The SMC signal consists of 2 components: continuous component u_{eq} and discontinuous component u_n whose relation is given by:

$$u = u_{eq} + u_n \quad (15)$$

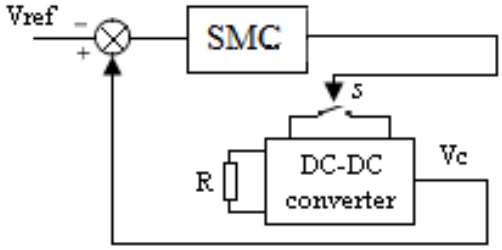


Fig. 2 : Block of SMC controlled Buck Converter.

Thus the existence of the sliding mode implies that $\dot{s} = 0$.

In the linear controller, though the rise time and settling time of the system are appropriate for system stability and magnitude of the overshoot is reduced to a certain extent, their weak response to the parameter variations decreases the system performance. But the nonlinear controllers hold good for variations in the system with better dynamic response and stabilizes the system completely with absence of overshoot [1].

C. Sliding Mode Proportional Integral Derivative Controller

For most of the converter applications, controllers with less complexity and higher efficiency are preferred, hence the performance of both the linear and non-linear controllers are combined together to stabilize the system efficiently at much faster rate. Thus the sliding mode PID voltage controller is widely used for its better performance and simple implementation[5]. The major drawback in the conventional SMC is the chattering problem, which is overcome by this Sliding Mode Proportional Integral Derivative controller accurately. Fig:3 represents the block diagram of SMCPID controlled buck converter.

The dynamic equation of SMC PID Controlled Buck Converter is:

$$\frac{dV_p}{dt} = k_4(x_2 - V_{ref}) \quad (16)$$

$$\frac{di_p}{dt} = \frac{1}{LC}\left[u - v_p + k_5(x_2 - V_{ref})\right] \quad (17)$$

The sliding mode surface is selected as:

$$s = \frac{x_2 - V_{ref}}{k_6/\sqrt{LC}} + LCi_p = 0 \quad (18)$$

Where k_4, k_5 *and* k_6 express sliding mode PID parameters.

V_{ref} is the reference voltage,

V_p is the peak voltage.

When s x □̇ < 0 is satisfied, the sliding mode surface is attractive. Once the system reaches the sliding mode surface, the following equations are obtained.

For ,

$$S = 0 \quad i_p = \frac{1}{k_6\sqrt{LC}}(x_2 - V_{ref}) \quad (19)$$

$$\dot{S} = 0 \quad u_{eq} = k_4 \int (x_2 - V_{ref})dt - \frac{\sqrt{LC}}{k_6}\frac{dx_2}{dt} - k_5(x_2 - V_{ref}) \quad (20)$$

The effect of the PID performance on the sliding surface depends on the gain parameters.

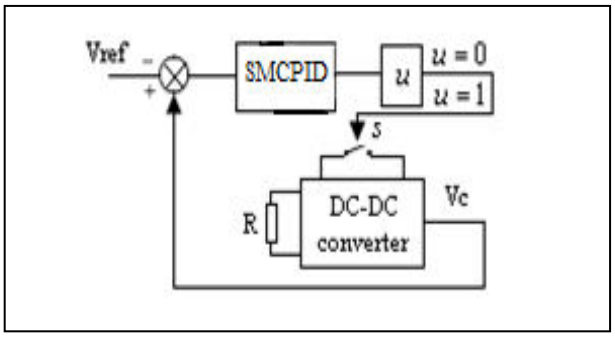


Fig..3 : Block of SMCPID controlled Buck Converter

When K_p increases, the front end oscillations are reduced, K_i reduces the output voltage oscillation and increase the system stability and K_d speeds up the system performance. Thus with proper tuning, the system output voltage reaches the reference voltage very fast and the system has good static and dynamic performance.

IV. SIMULATION AND DISCUSSION

The response of the buck converter in the open loop and closed loop condition for various gain parameters has been obtained using the MATLAB7.0 Simulink and shown in the following figures. From Fig:4 it is noted that maximum overshoot occurs in the system and hence to overcome this the conventional PID controller is enforced. Fig:5 shows the response of the PID controlled Buck converter. It can be seen that, though the maximum overshoot has been reduced to a certain extent, still the peak overshoot problem has to be overcome so the non-linear controller SMC is introduced. The SMC is tuned and the response for appropriate Sliding coefficient has been determined which is shown in Fig:6

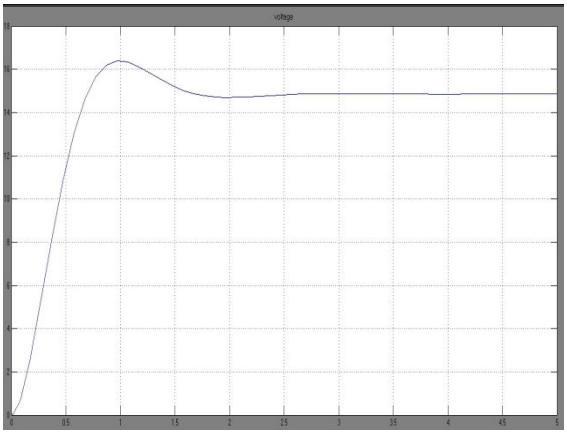


Fig. 4:Voltage Response of Open Loop Buck Converter

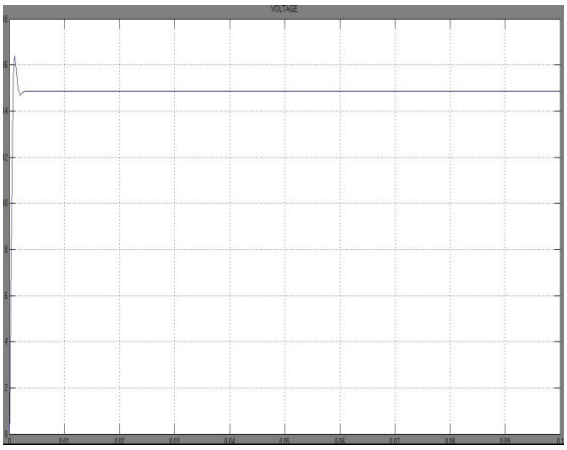


Fig.5 : Voltage Response of PID controlled Buck Converter

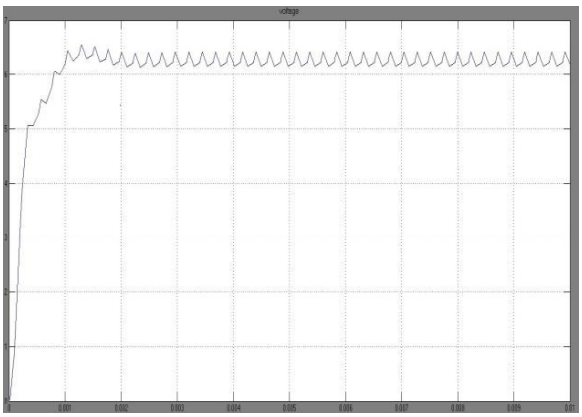


Fig.6: Voltage Response of SMC controlled Buck Converter.

From the response curve, it is evident that SMC gives a chattering effect which is nothing but the harmonics due to switching loss and electromagnetic interference. Therefore to overcome this, Sliding Mode Proportional Integral Derivative controller is used so that the system stabilizes efficiently and dynamically with no chattering effect. Fig:7 shows the response curve of the SMCPID controlled buck converter proving the effectiveness of SMCPID controller.

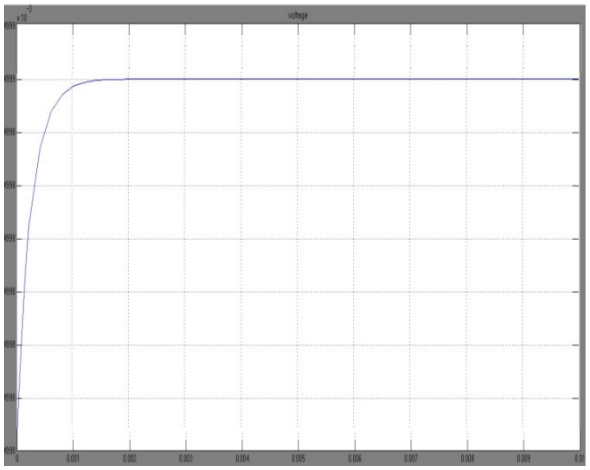


Fig. 7 : Voltage Response of the SMCPID controlled Buck Converter

From the response curves, the performance characteristics of the buck converter in the open and closed loop condition has been tabulated.Table 1 shows the performance charactersistics of the system.

Table 1: Performance Indices Of Open And Closed Loop Buck Converter

Performance Criteria	**Open Loop**	**Closed Loop**		
		PID	**SMC**	**SMCPID**
Peak Time(t_p)	0.00099	0.00091	-	-
Rise Time(t_r)	0.00063	0.00044	0.00033	0.0003
Delay Time(t_d)	0.00036	0.00036	0.00019	0.00007
Peak Overshoot(μ_p)	2.385v	1.53v	-	-
Settling Time(t_s)	0.00268	0.00265	0.0025	0.0020

Finally from the table also, it is proved that SMCPID controller gives a better performance comparatively.

V. CONCLUSION

The objective of analysing the efficiency and robustness of linear PID and non-linear SMC and SMCPID controller, for buck converter has been successfully done and the simulation results and performance indices are determined. From the results it is evidently proved that SMCPID is highly efficient, robust and dynamic in stabilizing the system over the other controllers. SMCPID also proves to be the best

solution to overcome the chattering problem in systems with load and switching frequency variations.

REFERENCE

[1] Kamel Ben Saad, Abdelaziz Sahbani and Mohamed Benrejeb, " Sliding Mode Control and Fuzzy Sliding Mode Control for DC-DC Converters," Research unit LARA, National Engineering School of Tunis (ENIT),Tunisia.

[2] Khalifa, Al-Hosani,Andrey Malinin and Vadim I.Utkin, "Sliding Mode PID Control of Buck Converters", The European Control Conference, 2009.

[3] Andrzej Bartoszewicz and Justyna Zuk, "Sliding Mode Control-Basic Concepts and Current Trends", Institute of Automatic Control, Poland, Invited paper,pp 3772-3777, 2010.

[4] Khalifa Al-Hosani, Vadim Utkin and Andrey Malinin, "Sliding Mode for Industrial Controllers" ,The Ohio State University, IKOR, USA.

[5] Hongmei Li and Xiao Ye, "Sliding Mode PID control of DC-DC Converter",5th IEEE conference on Industrial Electronics and Applications, pp 730-734, 2010.

[6] Utkin, V.I, "Variable structure systems with sliding modes", IEEE Transactions on Automatic Control, Vol.AC-22, No.2, pp.212-222, April 1977.

[7] Nguyan,V.M and Lee,C.Q, "Indirect Implementation of sliding mode control law in buck-type Conveters", Proc. Of IEEE Applied Electronics Conference and Exposition, vol.1 pp.111-115, , March 1996.

[8] Siew-Chong Tan, Lai and Chi K.Tse, "An evaluation of the practicality of Sliding Mode Controllers in DC-DC converters and their general Design Issues", Proc. Of the IEEE Power Electronics,vol.7 No.2,pp.187-193,June 2006.

[9] Geethanjali,P, Vijaya Priya,P, Kowsalya,M , Raju,J , "Design and Simulation of Digital PID Controller for Open loop and Closed Loop Control of Buck Converter", International Journal of Computer Science and Technology, vol .1,issue 2, pp 203-206, December 2010.

[10] Tsai,J.F and Chen,Y.P, "Sliding Mode control and Stability analysis of Buck DC-DC converter", International Journal of Electronics, vol.94,no.3, pp.209-222, March 2007.

[11] Muhammad H.Rashid. "Power Electronics Circuits , Devices and Applications", Prentice – Hall of India, New Delhi.

◈◈◈

Speech Recognition using SOM and Actuation via Network in Matlab

Subash Chandar. A[1], Suriyanarayanan. S[2] & Manikandan. M[3]

[1]National University of Singapore, Kent Ridge, Singapore, India
[2]Department of Electrical Engineering, Indian Institute of Technology, Chennai, India
[3]Converteam EDC Private Ltd., Chennai, India
E-mail : subash_394pec@yahoo.co.in[1], ussuriya@gmail.com[2], m.manibala@gmail.com[3]

Abstract - This paper proposes a method of Speech recognition using SOM and actuation through network in Matlab. The different words spoken by the user at client end are captured and filtered using LMS algorithm to remove the acoustic noise. FFT is taken for the filtered voice signal. The voice spectrum is recognized using trained SOM and appropriate label is sent to server PC. The client and the server communication are established using UDP. Microcontroller (AT89S52) is used to control the speed of the actuator depending upon the input it receives from the client. Real-time working of the prototype system has been verified with successful speech recognition, transmission, reception and actuation via network.

Key words - *User Datagram Protocol; Self organizing map; Fast Fourier transform; Actuation; Speech recognition; Least mean square.*

I. INTRODUCTION

The area of speech processing is developing and shows tremendous potentialities for widespread use in the future. A speech recognition system makes human interaction with computers possible through a voice/speech to initiate an automated service or process [1]. Controlling a machine by simply talking to it gives the advantage of hands-free, eyes-free interaction. Several literatures have been published for Speech recognition using neural networks [3]-[6]. Constructing an effective Speech recognition system requires an in-depth understanding of both the tasks to be performed, as well as the target audience who will use the final system. Actuation based on network offers unique advantage over traditional local control. Combining speech recognition with network actuation can be used to control the actuator from a remote place.

The system (Speech Recognition and Actuation via Network) is divided into two modules. The first module is the client module with speech recognition system which provides the interaction between the user and the PC. The words spoken by the user are captured by the client computer. The server module with serial communication is another one which receives data from the client computer using UDP. The server module provides the PC communication with the actuator through a micro controller. The client module is developed in M-script which in turn calls a Simulink file responsible for sending data to client. The server module is a Simulink file which receives data and sends it to serial port.

UDP is used for the network communication between host and remote computers. The server performs two functions such as data reception and data delivery. Data is received by the server from the client, which delivers a label value for each word captured. The data is used by the server to actuate any device using a microcontroller. The micro controller produces the necessary control signal with respect to the data received and it is sent to the actuator. The data from the server is transmitted via a serial port to the micro controller.

II. SELF ORGANISING MAP

A self-organizing map (SOM) is a type of artificial neural network which is trained using unsupervised learning. Self-organizing maps are different from other artificial neural networks in the sense that they use a neighborhood function to preserve the topological properties of the input space. The principle goal of the self-organizing map is to transform an incoming signal pattern of arbitrary dimension into a one- or two dimensional discrete feature maps, and to perform this transformation adaptively in a topologically ordered fashion.

During training phase, the input vector from training set is fed to the network, its Euclidean distance to all weight vectors is computed. The neuron with weight vector most similar to the input is called the winning neuron **w**. The weights of the winning neuron and neurons close to it in the SOM lattice are adjusted towards the input vector. The size of the topological neighborhood function shrinks with time because of time varying width, hence the distance from the winning neuron. The update formula for a neuron with weight vector is given by (1), (2), (3) and (4).

$$w_j(n+1) = w_j(n) + \eta(n)h_{j,i(x)}(n)\left(x - w_j(n)\right) \quad (1)$$

$$h_{j,i(x)}(n) = e - \left(\frac{d^2{}_{j,i}}{2\sigma(n)^2}\right) \quad (2)$$

$$\sigma(n) = \sigma_0 e - (\tfrac{n}{\tau 1}) \quad (3)$$

Where h is the topological neighborhood function, d is the Euclidian distance from the neuron j to the winning neuron i, σ_0 and σ is the initial and time varying effective width of the topological neighborhood and η is the learning weight, n is number of iteration count value. τ_1 is the time-constant to control the decay rate of the learning rate which is given by,

$$\tau 1 = \frac{T}{log(\sigma_0)} \quad (4)$$

The new weight is the average of the input and the current weight. The synaptic weight vector $\mathbf{w}_i$ of winning neuron i move toward the input vector **x**. All the neurons in the neighborhood of the winning neuron also move toward the input vector. The farther away neurons have less change. Upon repeated presentations of the training data, the synaptic weight vector tends to follow the distribution of the input vectors due to the neighborhood updating. Thus leading to a topological ordering of the feature map in the sense that neurons that are adjacent in the lattice as shown in Fig.1, will tend to have similar synaptic weight vectors.

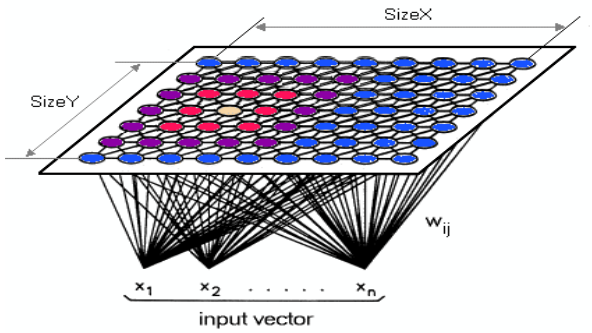


Fig.1 : Lattice Structure of SOM

III. SYSTEM ARCHITECTURE

The architecture is very simple and it comprises the following systems working in a combined manner, Speech Recognition system, Client PC, Network, Server PC, Serial port Interface and an Actuator (Motor).

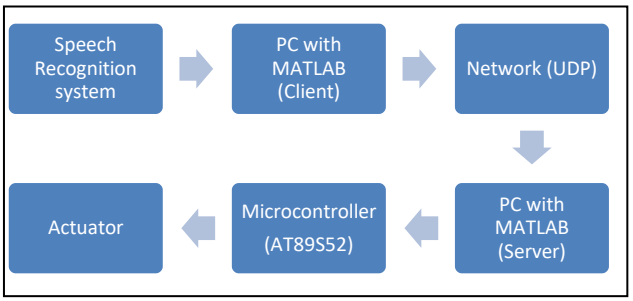


Fig. 2: System architecture of Network based Actuation

A. *Speech Recognition System*

The Speech Recognition system consists of speech acquisition module and recognition module. The speech acquisition module records the user voice for two seconds preceded by a beep sound which indicates the acquisition of data and saves it as a row vector as shown in Fig.3. The raw voice data is filtered using LMS algorithm to remove the acoustic noise Fig.4. FFT is taken for filtered voice signal to reduce the vector length hence reducing the execution time. The speech recognition module is a trained SOM, which gives a label to the acquired data as per the trained value. The training data for SOM has four words from the user which are, 'START','STOP','SPEED1','SPEED2' and 'BREAK'. The training dataset is formed by taking FFT [2] of the input voice as shown in Fig.5, to avoid the difference due to delay of speech. Another advantage of taking FFT is, the spectrum remains same for region of interest, irrespective of cross interference and noise. The dataset consists of voice signals of different individuals both male and female. The trained network is then used to identify four words which are specified above. The cluster formation for different words spoke by different individuals by SOM is shown in Fig.6.

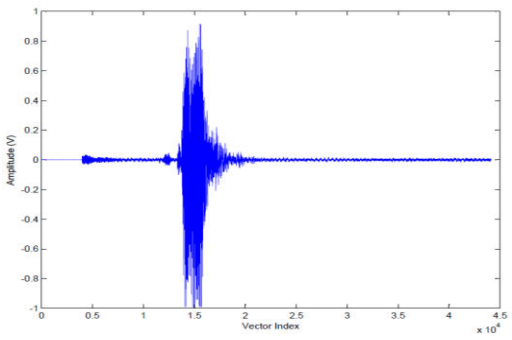


Fig. 3 : Raw voice signal

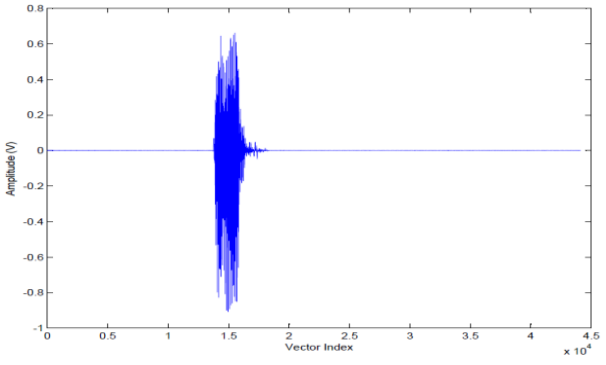


Fig. 4 : LMS filtered voice signal

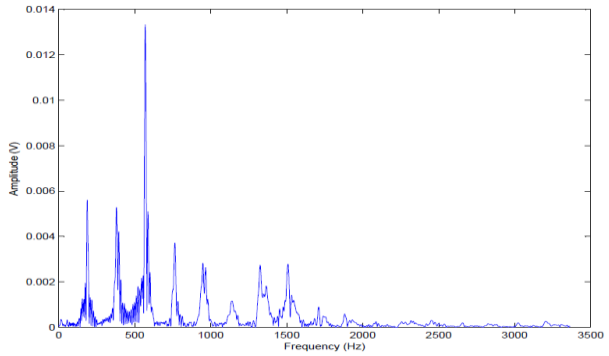


Fig. 5 : FFT of filtered of voice signal

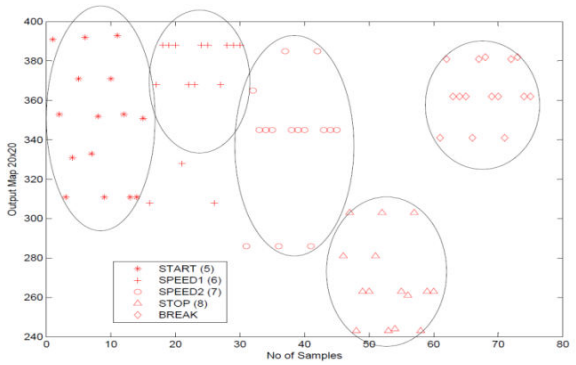


Fig. 6 : SOM mapping for different words

B. *Client (UDP send)*

The client module is designed in Simulink. It is developed to receive the data from Speech Recognition module and to move the received data to the server via network using UDP in real time, with continuous polling. The port address of server, dynamic IP address and local port for binding is given as input to the client module. The UDP communication tool is taken from Instrument Control Toolbox in Simulink. The data transmitted to server is a 1 byte character, label value for each of four words. The Word and their corresponding label values are shown in Table 1.

Table 1 : Word and its Corresponding Values

WORD	START	SPEED1	SPEED2	STOP	BREAK
SOM LABEL	5	6	7	8	Exit
PULSE FREQUENCY	48Hz	24Hz	12Hz	0Hz	Exit

C. *Network*

The protocol used for network communication is User Datagram Protocol. The UDP is one of the core protocols of the Internet protocol suite. UDP provides communication service at an intermediate level between the client PC and the Internet protocol. The unassigned port in UDP ranges from 0 to 65535. The communication port and local binding port address used in our prototype module is 45000.

D. *Server (UDP receive and Serial send)*

The server module is a Simulink file. It performs two different functions in real time.

1) *UDP receive:* The server Simulink should be run before executing the client module. If this condition fails, the client will automatically quit, prompting "timeout error". The server will be in continuous polling and will check for any data in the specified port. If the data appears in the port of declared IP address of client, it passes the data to data transmission module.

2) *Serial send:* This block configures the serial port with the specified parameter (baud rate: 9600, Data bits: 8, Stop bit: 1, No parity) and sends the data once it gets from the UDP receive module.

E. *Serial Port Interface*

As most of the current day PCs are lacking serial port, A USB to RS232 converter is used for serial port communication which can be configured to any unused port to avoid hardware resource sharing.

F. *Microcontroller-AT89S52*

Microcontroller is a microprocessor designed specifically for control applications, and is equipped with ROM, RAM and facilities I / O on a single chip. AT89S52 is one of the family MCS-51/52 equipped with an internal 8 Kbyte Flash EPROM (Erasable and Programmable Read Only Memory), which allows memory to be reprogrammed. There are four input and output ports each consist of 8 bits. The clock frequency used is 12 MHz. The program is done in assembly language and burnt using KEIL C.

The data from the serial port is fed into SCON (Serial Communication) pin of microcontroller (AT89S52). The microcontroller decodes the value and

saves it in the accumulator. Depending on the accumulator value different frequency is generated. The generated pulse is fed via port 2, which is connected to a low power DC motor. Thus for different values received from the client the speed of the motor will vary. The actuator can be directly driven by a microcontroller, if the power rating is less else a driver circuit has to be used to avoid loading from actuator.

IV. WORKING PRINCIPLE

The flow of data is explained in Fig.2. The speech recognition system captures the user voice in real time and recognizes the word spoken by the user after subtracting the ambient noise, with the help of trained SOM. The UDP send block is called by the Speech Recognition module which passes the appropriate label value mentioned in Table 1 to UDP receive block. The UDP receive block receives the data from UDP send block and sends it to Serial send block. The serial configuration block configures the serial port. The serial send block pushes the data received from the client to the Micro-controller. A look up table is developed in the microcontroller which generates frequency appropriate to the data received from the server. The speed of the actuator is controlled by the words, 'SPEED1' and 'SPEED2'.

V. HARDWARE SET UP AND EXPERIMENTAL RESULTS

A. *Hardware Setup*

To verify the architecture and working principle, a hardware prototype as show in Fig.7, is built using Microcontroller AT89S52. The code is developed in M-Script and a model is formed using Simulink in Matlab.

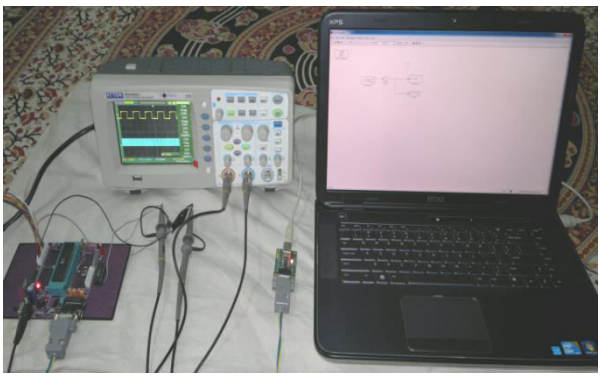
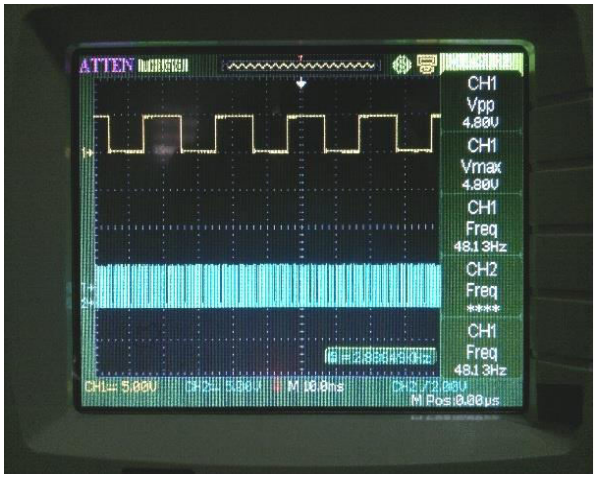


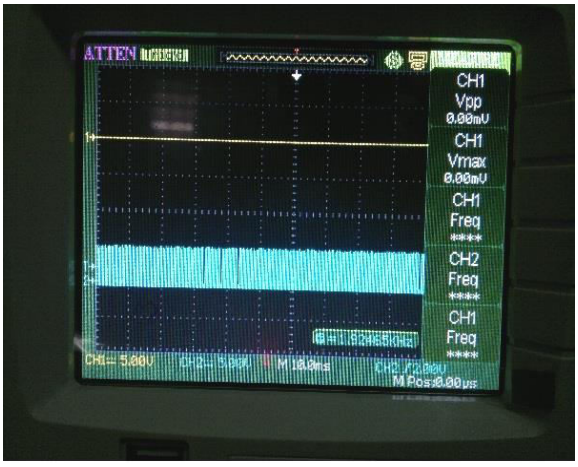
Fig.7 : Hardware set up showing Server PC running Simulink model with Microcontroller and DSO waveforms

The communication via network is done using wireless router which assigns IP address dynamically to the PCs connected to the network. USB to RS232 is used in the hardware set up for serial communication with the microcontroller. A microcontroller with serial port interface is used for decoding of serial data from server. The microcontroller is programmed to receive the data from server at a baud rate of 9600bps. The pulse is generated at port 2 of the microcontroller. A DSO is used to check the pulse pattern. Pulse of frequency 48, 24, 12 and 0 HZ is generated for the data received from the client in real time. With help of these pulses, a motor can be actuated for different speed. The pulse waveform for 'START' and 'STOP' word are shown in the Fig.8 (a) and Fig.8 (b) respectively.

B. *Experimental Results*



a. Received Value (8 Bit) =5, Frequency =48Hz



b. Received Value (8 Bit) =8, Frequency =0Hz

Fig. 8 : Pulses from server for START and STOP word.

VI. CONCLUSION

Speech is captured and recognized in real-time with help of SOM developed in M-script and transmitted by UDP send block from Instrumentation and control

toolbox of Simulink. The obtained real time raw data is converted into frequency spectrum before analysis to increase the accuracy of SOM. transmitted. The server module is able to receive the data by UDP receive block and send it to Microcontroller (AT89S52) with serial send block. 'Enable Blocking Mode' in serial send block of Simulink model is disabled to reduce time delays between data transmission. Data size parameter in UDP receive block should be declared appropriately to avoid fluctuation in the pulse. Pulse of appropriate frequencies is generated for unique label value received from the trained SOM, which can be used for actuation. A lab prototype is built and all the experiments are conducted in real time. The results are obtained and found to be satisfactory.

VII. FUTURE WORK

The simplex communication method used in this paper can be done using Duplex communication through which real human-computer interaction can be visualized. To achieve more smoothing actuation, the no of words can be increased in the proposed topology can be increased. The SOM can be trained for different ascent and its performance can be evaluated. With some hardware modification, the proposed method can be used for Voice controlled home automation.

REFERENCES

[1] Punit Kumar Sharma, Dr. B.R. Lakshmikantha and K. Shanmukha Sundar, "Real Time Control of DC Motor Drive using Speech Recognition" , Power Electronics (IICPE), 2010 India International Conference.

[2] Shlomo Engelberg, Yishai Saidoff, and Yehezkel Israeli, "Voice Identification Through Spectral Analysis", Instrumentation & Measurement Magazine, IEEE Volume: 9, Issue: 5.

[3] Chenghui Yang, Weixin Yang and Shuwen Wang, "Based on Artificial Neural Networks for Voice Recognition Word Segment", Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference.

[4] Akshay Gupta, "Synthesis and Performance Analysis of a Recurrent Fuzzy Multilayer Perceptron for Speech Recognition", Methods and Models in Computer Science (ICM2CS), 2010 International Conference.

[5] TAbushariah, A.A.M.,Gunawan, T.S., Khalifa, O.O. and Abushariah, M.A.M., "English Digits Speech Recognition System Based on Hidden Markov Models", Computer and Communication Engineering (ICCCE), 2010 International Conference.

[6] Hasnain, S.K and Awan, M.S., "Recognizing Spoken Urdu Numbers Using Fourier Descriptor and Neural Networks with Matlab", Electrical Engineering. ICEE 2008, Second International Conference.

◈◈◈

Information Hiding In Color Images

R. S. Mahamune & P. J. Suryawanshi

Electronics and Telecommunication Dept., Priyadarshini College of Engineering, Nagpur, India
E-mail : rupesh.mahamune@rediffmail.com.com, pradnyajs@rediffmail.com

Abstract - Digital Watermarking is the technique for embedding information into a digital signal. The watermarking technique provides a persistent link between the authenticator and the content it authenticates. The digital image watermarking can be categorized into different category-Visible and Invisible as well as Fragile, Semi-fragile and Robust. In this kind of watermarking the information is added as digital data to the original, but it can not be perceived in Human Visual System (HVS). Image authentication has achieved a sharp attention now a day due to broad availability of Internet services. Mal distribution and illegal copying of image, volatiles the authenticity of image ownership. The proposed technique will ensure the complete authentication of ownership in a dynamic manner. The proposed technique forms the watermark from the host image itself by combining the unique zone of the host image, and then it is embedded to the host image with LSB scheme. It allows a user with an appropriate secret key and a hash function to verify the authenticity, integrity and ownership of an image. If a forger performs the watermark extraction with an incorrect key and inappropriate hash function, the user obtains an image that resembles noise. This provides integrated solution for ownership authentication where the watermark is unique for that particular host image, thus the authentication is ensured in an efficient way. At the watermark extraction end, blind extraction method is used, i.e., neither the host image nor the watermark image is required at the time of watermark extraction. The PSNR is a measure of quality of watermarked image. PSNR is provided only to give us a rough approximation of the quality of the watermark for different color images.

Key words - *authentication; ownership; invisible watermarking; LSB scheme.*

I. INTRODUCTION

Digital Watermarking is the technique for embedding information into a digital signal. The watermarking technique provides a persistent link between the authenticator and the content it authenticates. The digital image watermarking can be categorized into different category- Visible and Invisible as well as Fragile, Semi-fragile and Robust. In invisible fragile watermarking the information is added as digital data to the original, but it can not be perceived in Human Visual System(HVS). Suppose a user doesn't have a watermark or the user doesn't have that much of time to create the watermark. As a remedy to the aforesaid problem we've introduced a new framework that dynamically forms the watermark from the host image by combining the unique zone of the host image. As an added flavor the ownership authentication is guaranteed in an efficient manner. Because every host image will have its own dynamically formed watermark. The dynamic watermarking framework will ensure:

- *No watermark construction overhead.*
- *Uniqueness in dynamically formed watermark.*

II. TECHNIQUES OF WATERMARKING

A. Least Significant Bit Modification

The most straight-forward method of watermark embedding, would be to embed the watermark into the least-significant-bits of the cover object .Given the extraordinarily high channel capacity of using the entire cover for transmission in this method, a smaller object may be embedded multiple times. Even if most of these are lost due to attacks, a single surviving watermark would be considered a success. LSB substitution however despite its simplicity brings a host of drawbacks. Although it may survive transformations such as cropping, any addition of noise or lossy compression is likely to defeat the watermark. An even better attack would be to simply set the LSB bits of each pixel to one…fully defeating the watermark with negligible impact on the cover object. Furthermore, once the algorithm is discovered, the embedded watermark could be easily modified by an intermediate party. An improvement on basic LSB substitution would be to use a pseudo-random number generator to determine the pixels to be used for embedding based on a given "seed"

or key . Security of the watermark would be improved as the watermark could no longer be easily viewed by intermediate parties. The algorithm however would still be vulnerable to replacing the LSB's with a constant. Even in locations that were not used for watermarking bits, the impact of the substitution on the cover image would be negligible. LSB modification proves to be a simple and fairly powerful tool for stenography, however lacks the basic robustness that watermarking applications require.

B. Frequency Domain Techniques

An advantage of the spatial techniques discussed above is that they can be easily applied to any image, regardless of subsequent processing (whether they survive this processing however is a different matter entirely). A possible disadvantage of spatial techniques is they do not allow for the exploitation of this subsequent processing in order to increase the robustness of the watermark. In addition to this, adaptive watermarking techniques are a bit more difficult in the spatial domain. Both the robustness and quality of the watermark could be improved if the properties of the cover image could similarly be exploited. For instance, it is generally preferable to hide watermarking information in noisy regions and edges of images, rather then in smoother regions. The benefit is two-fold; Degradation in smoother regions of an image is more noticeable to the HVS, and becomes a prime target for lossy compression schemes. Taking these aspects into consideration, working in a frequency domain of some sort becomes very attractive. The classic and still most popular domain for image processing is that of the Discrete-Cosine-Transform, or DCT. The DCT allows an image to be broken up into different frequency bands, making it much easier to embed watermarking information into the middle frequency bands of an image. The middle frequency bands are chosen such that they have minimize they avoid the most visual important parts of the image (low frequencies) without over-exposing themselves to removal through compression and noise attacks (high frequencies) .

One such technique utilizes the comparison of middle-band DCT coefficients to encode a single bit into a DCT block. To begin, we define the middle-band frequencies ($F_{M)}$ of an 8x8 DCT block as shown below in figure 1.

F_L is used to denote the lowest frequency components of the block, while F_H is used to denote the higher frequency components.

F_M is chosen as the embedding region as to provide additional resistance to lossy compression techniques, while avoiding significant modification of the cover image.

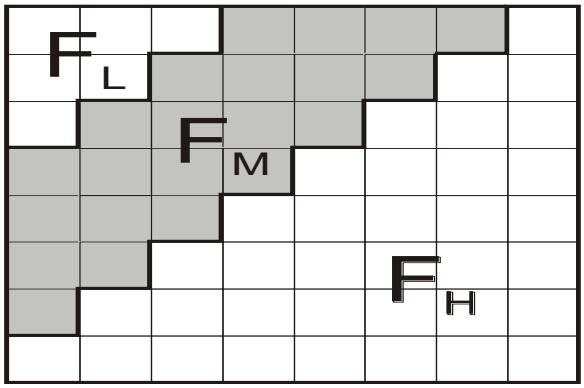


Fig.1 : Definition of DCT Regions

Next two locations $B_i(u_1,v_1)$ and $B_i(u_2,v_{2)}$ are chosen from the F_M region for comparison. Rather then arbitrarily choosing these locations, extra robustness to compression can be achieved if we base the choice of coefficients on the recommended JPEG quantization values. If two locations are chosen such that they have identical quantization values, we can feel confident that any scaling of one coefficient will scale the other by the same factor…preserving their relative size.

III. TECHNIQUE FOR FORMATION OF WATERMARK

Convert the host image from RGB model to YST model. Then we divide the host image in 'n' number of different blocks depending on the size of the image. After that we apply Discrete Cosine Transform (DCT) on such blocks and identify the maximum pixel from all blocks those contain the most high frequencies. Combining those pixels we form the watermark image dynamically.

A. Algorithm for formation of watermark

- Convert the Host Image from RGB model to YST model.
- Divide the host image in 'n' number of different blocks depending on the size of the image.
- After that apply Discrete Cosine Transform (DCT) on such blocks.
- Identify the maximum pixel from all blocks those contain the most high frequencies.
- Combine those pixels we form the watermark image dynamically.

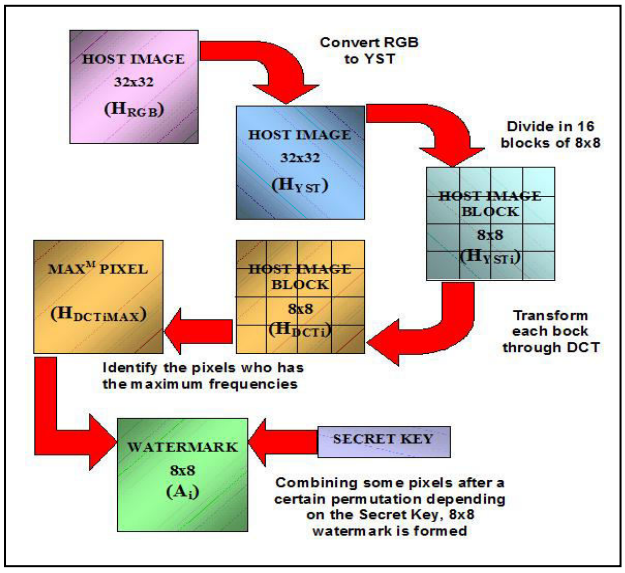


Fig. 2 : Schematic Diagram to form Watermark

B. Execution of Algorithm for formation of watermark

$$H_i = \begin{bmatrix} 117\$116\$111 & 117\$116\$111 \ldots . & 115\$118\$111 & 116\$117\$111 \\ 121\$120\$115 & 120\$119\$114 \ldots . & 110\$113\$106 & 109\$112\$105 \\ . & & . & \\ . & & . & \\ . & & . & \\ . & & . & \\ 111\$112\$106 & 111\$112\$106 \ldots . & 104\$109\$102 & 106\$109\$102 \\ 111\$110\$105 & 110\$111\$105 \ldots . & 105\$110\$103 & 107\$110\$103 \end{bmatrix}$$

- *Input: Host Image Block*
- Convert Hi from RGB color model to YST color model to form HYSTi.

$$H_{YSTi} = \begin{bmatrix} 115.75 & 115.75 \ldots . & 116.32 & 119.75 \\ 118.75 & 118.03 \ldots . & 119.04 & 118.03 \\ . & & & . \\ . & & & . \\ . & & & . \\ . & & & . \\ 106.73 & 107.33 \ldots . & 109.32 & 108.72 \\ 107.72 & 108.33 \ldots . & 108.72 & 107.72 \end{bmatrix}$$

- Transform HYSTi through Discrete Cosine Transform to obtain HDCTi.
- Identify all such HDCTi pixels those contain most high frequencies.
- Combining those pixels after certain permutation depending on a secret key the watermark A is formed.

IV. TECHNIQUE FOR EMBEDDING WATERMARK AND EXTRACTING THE WATERMARK

Extract LSB's of 'Blue (B)' value from each pixel of individual blocks of host image using a secret key and a hash function. 'Blue' is the most high frequency part of an image. That's why it is less sensitive to Human Visual System (HVS). Only the 'Blue' value of the color host image is to be manipulated.

A. Algorithm for Embedding Watermark using LSB Scheme

Input: Color host image block (Yi), Watermark image(A), Secret key (K).

Output: Watermark embedded image block (Wi).

- Get the pixel information from the host image (Yi) to get (Yip).
- The LSB of the binary values of Yip is converted to zero to form Li.
- Generate Hi depending on the value of Li and The 16-bit secret key (K).
- Generate Ai by extracting LSB of blue value of each pixel of A.
- Generate Xi by XORing Hi and Ai.
- Generate Wi by replacing the LSB of Yip with Xi.

B. Algorithm for Extracting Watermark using 1-bit Scheme:

Input: Watermark embedded image block (Wi), Secret key (K).

Output: Extracted watermark image block (A_i).

- The LSB of the binary values of Wi is converted to zero to form L_i.
- Hi is generated depending on the value of L'i and The 16- bit secret key (K).
- Ri formed by extracting LSB of blue value of each pixel of Wi.
- A_i is formed by XORing Hi and Ri.

C. Execution of Algorithm for Embedding Watermark using

LSB Scheme:

Input: Host Image block

$$Y_i = \begin{pmatrix} 155\$153\$154 & 153\$153\$153 & 154\$154\$154 & 156\$158\$157 \\ 155\$153\$154 & 54\$154\$154 & 55\$155\$155 & 156\$158\$157 \\ 153\$151\$152 & 154\$154\$154 & 156\$156\$156 & 156\$158\$157 \\ 152\$150\$151 & 154\$154\$154 & 157\$157\$157 & 156\$158\$157 \end{pmatrix}$$

Watermark Image:

$$A = \begin{pmatrix} 172\$197\$168 & 154\$183\$153 & 146\$180\$145 & 144\$178\$143 \\ 155\$186\$155 & 142\$173\$141 & 137\$171\$134 & 139\$171\$134 \\ 144\$180\$144 & 137\$171\$134 & 141\$173\$134 & 140\$170\$132 \\ 145\$179\$142 & 141\$173\$134 & 141\$172\$131 & 138\$169\$128 \end{pmatrix}$$

Secret Key: **K= 1111 0000 1010 1100**

The binary information of blue value of a pixel of Yi is given by Yip.

$$Y_{ip} = \begin{pmatrix} 10011010 & 10011001 & 10011010 & 10011101 \\ 10011010 & 10011010 & 10011011 & 10011101 \\ 10011000 & 10011010 & 10011100 & 10011101 \\ 10010111 & 10011010 & 10011101 & 10011101 \end{pmatrix}$$

The binary information of blue value of a pixel of A is given by Ab.

$$A_b = \begin{pmatrix} 10101000 & 10011001 & 10010001 & 10001111 \\ 10011011 & 10001101 & 10000110 & 10000110 \\ 10010000 & 10000110 & 10000110 & 10000100 \\ 10001110 & 10000110 & 10000011 & 10000000 \end{pmatrix}$$

The LSB of the binary values of Yip is converted to zero to form Li.

$$L_i = \begin{pmatrix} 10011010 & 10011000 & 10011010 & 10011100 \\ 10011010 & 10011010 & 10011010 & 10011100 \\ 10011000 & 10011010 & 10011100 & 10011100 \\ 10010110 & 10011010 & 10011100 & 10011100 \end{pmatrix}$$

Hi is generated depending on the value of Li and The 16-bit secret key (K).

Ai is formed by extracting LSB of blue value of each pixel of A.

$$A_i = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Xi is formed by XORing Hi and Ai.

$$X_i = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Wi is formed by replacing the LSB of Yip with Xi.

$$W_i = \begin{pmatrix} 10011010 & 10011000 & 10011010 & 10011101 \\ 10011010 & 10011011 & 10011010 & 10011101 \\ 10011000 & 10011011 & 10011100 & 10011101 \\ 10010110 & 10011010 & 10011100 & 10011100 \end{pmatrix}$$

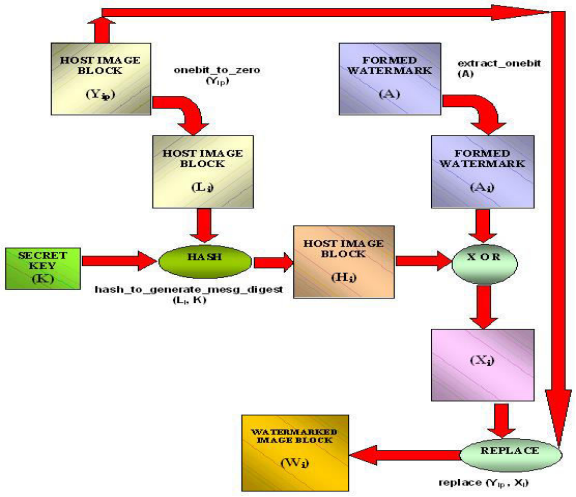


Fig. 3 : Schematic Diagram for Embedding Watermark

Fig. 3.a : Host Image

Fig.3.b:Formed Watermark

Fig.4.a:Watermarked Image

Fig.4.b:Extracted Watermark

REFERENCES

[1] Pradosh Bandyopadhyay, Soumik Das, Shauvik Paul, "A Dynamic Watermarking Scheme for Color Image Authentication" - International Conference on Advances in

[2] Soumik Das, Pradosh Bandyopadhyay, Shauvik Paul, Arindam Sinha Ray, Dr. Monalisa Banerjee, "Pioneering the Technique for Invisible Image Watermarking on Color Image"— International Journal of Recent Trends in Engineering, Volume I, May, 2009.

[3] Soumik Das, Pradosh Bandyopadhyay, Shauvik Paul, Arindam Sinha Ray, Dr. Monalisa Banerjee, "A New Introduction towards Invisible Image Watermarking on Color Image,"— Published in Proceedings of IEEE International Advance Computing Conference (IACC'09), India, 2009 (ISBN: 978-1-4244-2927-1)

[4] M. Hamad Hassan, S.A.M. Gilani, "A Semi-Fragile Watermarking Scheme for Color Image Authentication"- World Academy of Science, Engineering and Technology 19, 2006.

◈◈◈

Nonlinear System Identification Using A Novel Immune Artificial Fish Swarm Algorithm

Ritesh Kumar & Nishant Sahay

School of Electrical Sciences, Indian Institute of Technology, Bhubaneswar, India
E-mail : riteshiitbbs@gmail.com, nishant@iitbbs.ac.in

Abstract - This paper proposes a functional link artificial neural network(FLANN) model trained using a modified fish swarm optimization (FSO) algorithm for nonlinear system identification. The system modelling problem has been reformulated as an optimization problem. The FSO algorithm has been modified by incorporating the immunity features of the artificial immune systems. Simulation study reveals improved performance of the proposed algorithm over the conventional FSO algorithm for nonlinear system identification.

Key words *- FLANN; Fish swarm optimization; Immune system.*

I. INTRODUCTION

Science is the study of nature to answer solutions in some problem domain and Computational algorithms and tools are all inspired by Nature to solve our real time problems in an efficient way. System identification plays an important role in many areas of research now-a-days. The linear static systems can be easily identified using traditional least mean square algorithm. However, in practice, most of the systems which we encounter are nonlinear ones. In such systems, the traditional approaches do not yield satisfactory results. Therefore, advanced nonlinear approaches have been developed for such systems. Fish swarm algorithm is basically the simulation of behaviour of school of fishes. Fish can find the area with more nutritional value through individual search or by following other fishes. This trait is emulated by fish swarm algorithm and put to use in several nonlinear identification problem.

Nonlinear System Identification is of prime importance nowadays. Its applications are manifold. It is being used in varied areas such as electrical engineering, molecular biology, interpretation of microarray data, identification of MIMO systems etc.

Nonlinear System Identification can be accomplished using various kinds of adaptive algorithms. Neural networks represent an important method in classification of patterns and approximating complex nonlinear systems. Due to these properties, neural networks are preferred models for nonlinear systems whose mathematical models are difficult to obtain. A wide variety of neural networks models have been applied for such approximation purposes and these include models like FLANN(functionally linked artificial neural network), LeNN(Legendre neural network), PPN(polynomial perceptron network) etc.

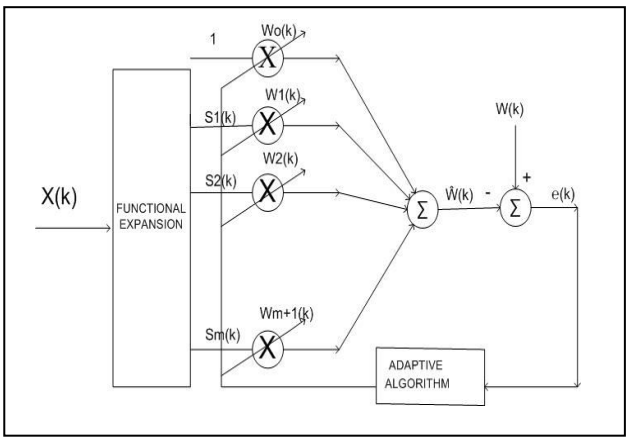


Fig. 1 : FLANN

FLANN was first proposed by Pao as a single layer ANN structure capable of forming arbitrarily wide complex Decision regions by generating non-linear decision boundaries.The use of FLANN not only increases the learning rate but also reduces the computational complexity.

The input signal X(k) is functionally expanded to a nonlinear values to an adaptive linear combiner whose weights are varied as per the iterative learning rule. Generally, LMS is used but here we have implemented Immune artificial fish swarm optimization algorithm to

change the weights. In this case,we have used trigonometric based linear expansion matrix given by:

$$\emptyset i\big(u(k)\big) = \begin{Bmatrix} 1\, for\, i = 0 \\ u(k) for\, i = 1 \\ \sin\big(i.pi.u(k)\big)\, for\, i = 2,4,\dots m \\ \cos\big(i.pi.u(k)\big)\, for\, i = 3,5,\dots .m+1 \end{Bmatrix}$$

Where i=1,2,….,m\2.As a result the total expanded values including an unity bias input becomes 2m+2.

Let the corresponding weight vector be represented as $W_i(k)$ having 2m+2 elements. The estimated output of the non-linear static part is given by-

$$F\big(u(k)\big) = \sum_{i=1}^{2m+2} Wi\emptyset i\big(u(k)\big) + \in (k)$$

Where, $\in (k)$ is approximation error.

From nature only we wanted to add one more element to our algorithm and that one was immunity.

Immunity is the ability of body to resist against diseases. The Clonal selection principle of AIS describes the activities of immune cells against the entry of pathogens or antigens and is simple and effective evolutionary computation tool to solve optimization problems.

The affinity of every cell with each other is a measure of similarity between them and is calculated by the distance between them. The antibodies present in memory has got higher preference than the one detected in primary response. During mutation, fitness as well as the affinity of antibodies gets changed.So,in each iteration antibodies with higher fitness and affinity are stored in memory and the low affinity cells are discarded.

The Clonal selection algorithm has several interesting features like adjustable search space, population size, location of multiple optima etc.

II. DEVELEOPMENT OF IAFSO

A. Artificial Fish Swarm Algorithm

As mentioned before, artificial fish swarm algorithm is the simulation of behaviour of a school of fishes, which find the zone with more nutritional value using individual search or by following other fishes. This algorithm comprises of three behaviours: preying, swarming and following behaviour. Preying behaviour is the most common behaviour for obtaining food. Fishes perceive the concentration of food through vision or sense to determine movement and then choose the tendency. Swarming behaviour is a kind of living behaviour in which the fishes assemble in groups naturally in moving processes to guarantee the existence of colony and to avoid danger. while swarming they obey three principles, which are namely Compartmentation principle, Unification principle and Cohesion principle. Compartmentation principle refers to avoiding congestion with other fellows. Unification principle refers to moving towards average fellows' moving direction whereas the Cohesion principle refers to the movement towards the centre of the group. Also, when a fish or a group of fishes find the food, the nearby fishes approach the food by following them as quickly as possible. This is known as the following behaviour. Before constructing the artificial fish(AF) model, we will introduce some definitions first.

Xi is the position of AFi. Y=f(X) is the fitness or objective function at position S, which can represent food concentration. Xij represents the distance between AFi and AFj . Visual represents the visual scope whereas 'a' represents the crowd factor of AF. 'mf' is the number of fishes within the visual scope of one fish. 'M' step is the maximum step of the moving AF and step is a random positive number within 'M' step. S(Xijk)<‖Xi −Xj ‖ < V isual) is the set of all the AF which lie within the visual scope of Xi exploring the area at the present position. The typical behaviour of AF can be expressed as follows:

Searching behaviour :In general, the fish stroll at random. When the fish(es) discover a water area with more food, they will go quickly towards the area. Let us assume that Xi is the AF state at present, and Xj . The beahviour can be expressed as follows:

$Xj = Xi + visual.rand()$

$Xi = Xi + step.(Xi - Xj)/\|Xi - Xj\|$ if $Yj > Yi$

$Xi^{(t+1)} = Xi^{(t)} + step.rand()$ else

Swarming behaviour :In the process of swimming, the fish will swarm spontaneously in order to share the food of the swarm. Let us assume that Xi is the AF state at the present and Xc is the centre of the swarm. Then, the fish behaviour can be expressed in formula as follows:

$$Xi^{(t+1)} = Xi^{(t)} + step.Rand().(Xc - Xi^{(t)})/\|Xc - Xi^{(t)}\|$$

Following behaviour :When one fish of the fish swarm discovers more food, the other fish will follow it. Let us assume that Xj is the position of the fish which has discovered food and Xi is the general state of AFi. Then the behaviour can be expressed as follows:

$$Xi^{(t+1)} = Xi^{(t)} + step.Rand().(Xj - Xi^{(t)})/\|Xcj - Xi^{(t)}\|$$

B. Immune Artificial Fish Swarm Algorithm

The Immune Artificial Fish Swarm algorithm is an improvement over the general one. In this approach, we

clone those fishes which are very near to the food source and clone them so that the population searching for better solution converges to it fastly and does not get stuck in any local solution.

Since ,in Fish swarm optimization,Swarm best position is the best position in the available search space,So each fish instead of following its personal best value will now tend to follow the swarm best value to reach to food faster with added Immunity to it to prevent it from straying away from solution space .

As per clonal selection principle when an antigen or a pathogen invades the organism, numbers of antibodies are produced by the immune cells. The fittest antibody undergoes cloning operation to produce number of new cells. These are used to eliminate the invading antigens.

Employing this principle of AIS in FSO we propose thateach particle is led to the Swarm best position wherefrom the next search is started.

Mutation operation is also implemented to diversify the search space.The Fitness value are evaluated then for the Mutated fish as well as the Fish present in the visual scope and then the best fishes are selected for the journey of traversing.

III. IAFSO BASED LEARNING

1. Generate 'K' number of input-output patterns which are required to learn the network uniformly distributed.
2. Each Input pattern 'X' is functionally expanded.
3. Each of the desired output is compared with corresponding desired output to generate 'k' errors and hence mean squre error for a parameter is determined by

$$MSE(n) = \sum_{i=1}^{k} (e(i)^2)/k$$

4. Weights are varied by using AFSO based learning mechanism.
5. The MSE is plotted to obtain the learning characteristics.Learnng is stopped when minimum MSE levels are reached or maximum number of iterations are achieved.

IV. SIMULATION STUDY

In this section we carry out the simulation study of FLANN based artificial fish swarm algorithm. The block diagram of Fig.1 is simulated where the coefficients or the weights of the FLANN model are updated using LMS and AFSA. We have carried out simulation without noise as well as with noise of SNR 20dB.Simulations are carried out extensively for different sets of problems to test the functionality of new proposed algorithm using MATLAB.

The accuracy of identification of proposed model assessed by following results:

1. Comparing the response curve of Inputs and Outputs.
2. Comparing the MSE curve.

Now we have tested the workability of this novel algorithm on the following examples:

Example 1. : Linear system [0.26000.93000.2600]

Example 2. : Test function The impulse response of the plant is [0.2600, 0.9300, 0.2600] and nonlinearity associated is yn(k) = tanh(y(k)) SNR 20dB.

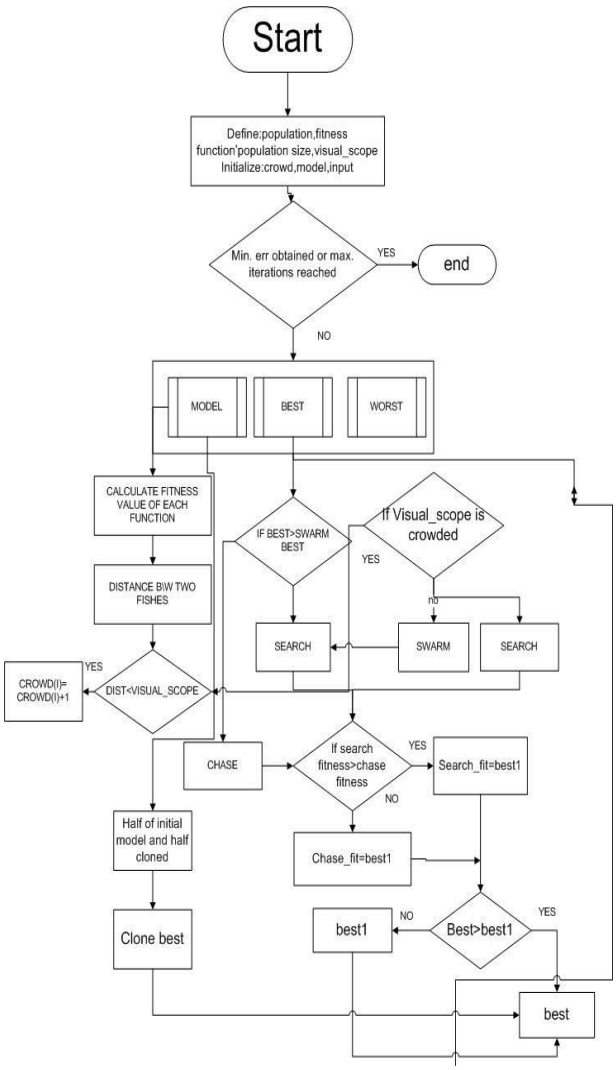


Fig. 2 : Flowchart

Example 3 : Test function Parameters of the linear system of the plant is [0.2600, 0.9300, 0.2600] and nonlinearity associated is

$yn(k) = y(k)+0.2[y(k)]^2-0.1[y(k)]^3$ SNR 20dB.

Where, y(k) is the output of the linear part of the plant plant and $y_n(k)$is the output of the overall system.

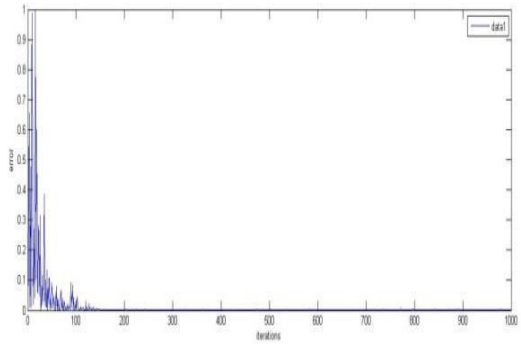


Fig. 1 : MSE curve for example 1

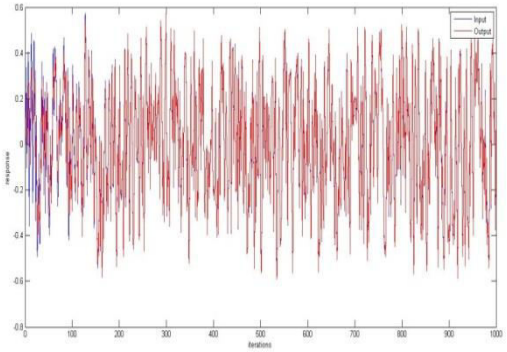


Fig. 2 : Response curve for example 1

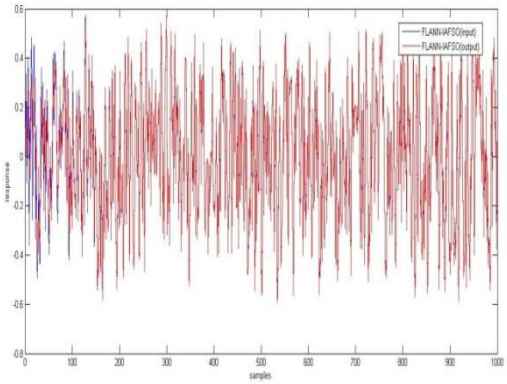


Fig. 3 : Response curve for example 2

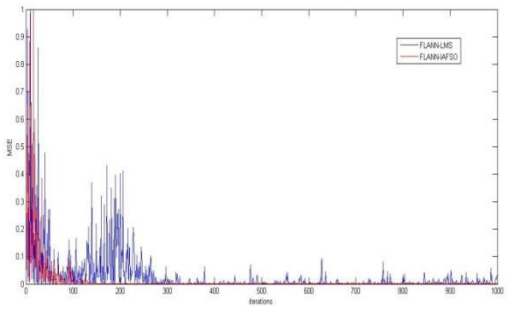


Fig. 4 : MSE curve for example 2

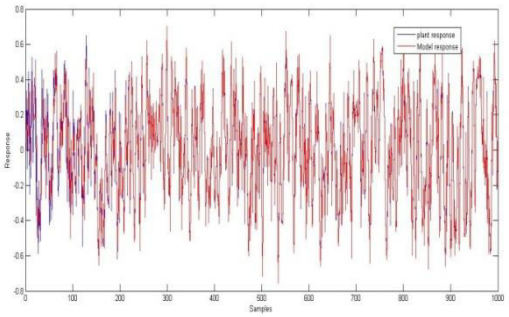


Fig. 5: response curve for example 3

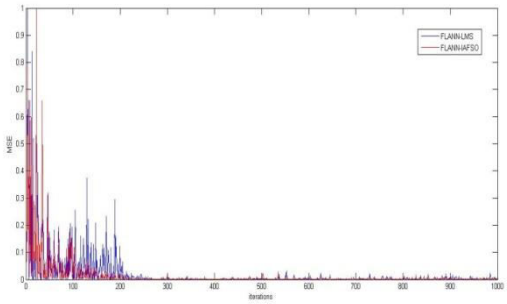


Fig. 6 : MSE curve for example 3

V. CONCLUSIONS

In this paper, we have modified the general fish swarm algorithm to convert it to immune fish swarm algorithm. We have used cloning which increases the convergence power of the solution and gives a more precise training result. In other words, we can say that the cloning of the fishes which give the best result provides immunity to the entire population. This process reduces the chance of straying of fishes from the best solution and also the chance of algorithm getting stuck in a local minimum. Thus this algorithm is both more convergent and best result giving as compared to its counterpart.

REFERENCES

[1] Yong Peng, An improved artificial fish swarm algorithm for optimal operation of cascade reservoirs, Journal of Computers, Volume 6,(April 2011).

[2] Hussein A. Abbass and Ruhul Sarker. Simultaneous Evolution of Architectures and Connection Weights in ANNs, in The Artificial Neural Networks and Expert Systems Conference (ANNES'01), Dunedin, New Zealand, pp. 16--21, November 2001.

[3] J.J.,1989a. AdaptiveIIRfiltering. IEEEASSP Magazine, 421. 1989, chapter 8,pp.197-222.

[4] Edite M. G. P. Fernandes1, Tiago F. M. C. Martins2 and Ana Maria A. C. Rocha1,FishSwarm Intelligent Algorithm for Bound Constrained Global Optimization,2009.

[5] Edite M. G. P. Fernandes1, Tiago F. M. C. Martins2 and Ana Maria A. C. Rocha1,Fish Swarm Intelligent Algorithm for Bound Constrained Global Optimization,2009.

[6] Yifen Hu, Parallel Fish Swarm Algorithm based on GPU Acceleration, Intelligent Systems and Applications.

[7] Mingyan jiang; Dongfeng yuan; Yongming cheng, Improved Artificial Fish Swarm Algorithm, in Natural Computation,2009.ICNC's09 Fifth International Conference.

[8] Salem F. Adra;Ian Griffin and Peter J.Fleming,Convergence Acceleration operator for mltiobjective optimization,IEEE transactions on evolutionary computation,Vol B,No.4,pp 825-847,August 2009.

[9] Babita majhi and G. Panda,"Cascaded Functional Link Artificial Neural Network Model for Nonlinear Dynamic System Identification ",International Journal of Artificial Intelligence and Soft Computing,vol.1,Nos.2/3/4,pp-223-237,2009.

◈◈◈

Design of Fault-Tolerant Memory with Fault Secure Encoder And Decoder

Sunitha R, Supritha G. S. & Ashwini

Bangalore
E-mail : suprithags@gmail.com, sunisathya@gmail.com

Abstract - Traditionally, memory cells were the only circuitry susceptible to transient faults. The supporting circuitries around the memory were assumed to be fault-free. Due to the decrease in feature size and power supply voltage, the rate of soft errors have increased and the logic circuits such as the encoder and decoder around the memory blocks have become susceptible to soft errors as well and must be protected. Memory cells have been protected from soft errors for more than a decade. Here, we introduce a reliable memory system that can tolerate multiple transient errors in the memory words as well as transient errors in the encoder and decoder (corrector) circuitry. Here we developed the definition of fault-secure detector (FSD) error correcting code (ECC) and the associated circuitry that can detect errors in the received encoded vector despite experiencing multiple transient faults in its circuitry. The structure of the detector is general enough that it can be used for any ECC that follows the FSD-ECC definition. We prove that Euclidean geometry low density parity check codes have the Fault secure detector capability. The proposed fault tolerant memory system is designed using VHDL. The implementation is for Spartan 3 FPGA using Xilinx ISE. Simulation results of the proposed fault tolerant memory system are shown using ModelSim and the simulation results are attached for the verification.

Key words - *ECC, Fault-secure detector, One-step majority logic corrector, Encoder, Decoder.*

I. INTRODUCTION

Due to the decrease in feature size and power supply voltage the modern VLSI integrated circuits are more and more susceptible to noise as well as transient faults caused by cosmic radiation(called single event upsets(SEUs)). As a result we can expect combinational logic to be susceptible to transient fault in addition to the storage cells and communication channels. Nowadays, not only charge particles from the sun but also neutrons have sufficient energy to flip bits in memories and corrupt logic inside processors. Therefore the paradigm of protecting only memory cells and assuming the surrounding circuitries (i.e., encoder and decoder) will never introduce errors is no longer valid. The major concern is logic, which, unlike memory cannot be easily protected using parity checking or Hamming error correcting codes (ECCs). Since the implementation of radiation hardened circuits is not always feasible or cost- efficient, designing highly-reliable digital systems using error detecting codes (EDCs) and implementing digital circuits as self-checking could be a viable choice.

One important class of digital circuitry, whose self checking implementation (to our knowledge) has not been considered yet, are encoders and decoders for ECC's used for data transmission. To date, in these applications, the major concern has been protection of data transmitted over the noisy channel and the data stored in memory cells rather than the protection of the process of encoding and decoding as well.

In this project, we introduce a fault-tolerant memory architecture which tolerates transient faults both in the storage units and in the supporting logic such as encoder, decoder (corrector), and detector circuitries. Particularly, we identify a class of error-correcting codes (ECCs) that guarantees the existence of a simple fault-tolerant detector design. This class satisfies a new, restricted definition for ECCs which guarantees that the ECC codeword has an appropriate redundancy structure such that it can detect multiple errors occurring in both the stored codeword in memory and the surrounding circuitries. We call this type of error-correcting Codes, fault-secure detector capable ECCs (FSD-ECC). The parity-check matrix of an FSD-ECC has a particular structure that the decoder circuit, generated from the parity-check Matrix, is Fault-Secure. The ECCs we identify in this class are close to optimal in rate and distance, suggesting we can achieve this property without sacrificing traditional ECC metrics. We use the fault-secure detection unit to design a fault-tolerant encoder and corrector by monitoring their outputs. If a detector detects an error in either of these units, that unit must repeat the operation to generate the correct output vector. Using this retry technique, we can correct potential transient errors in the encoder and corrector

outputs and provide a fully fault-tolerant memory system with fault tolerant supporting circuitry.

II. METHODOLOGY

In this section, we present our novel, restricted ECC definition for fault-secure detector capable codes.

The *minimum distance* of an ECC, d, is the minimum number of code bits that are different between any two codewords. The maximum number of errors that an ECC can detect is d-1 ,and the maximum number that it corrects is [d/2] . Any ECC is represented with a triple (n,k,d), representing code length, information bit length and minimum distance respectively.

For a particular ECC used for memory protection, let E be the maximum number of error bits that the code can correct and D be the maximum number of error bits that it can detect, and in one error combination that strikes the system, let e_e, e_m, and e_c be the number of errors in encoder, a memory word, and corrector and let e_{de} and e_{dc} be the number of errors in the two separate detectors monitoring the encoder and corrector units. In conventional designs, the system would guarantee error correction as long as $e_m \leq E$ and $e_e = e_c = 0$. In contrast, here we guarantee that the system can correct any error combination as long as $e_m \leq E$, $e_e + e_{de} \leq D$, and $e_m + e_c + e_{dc} \leq D$. This design is feasible when the following two fundamental properties are satisfied:

1) Any single error in the encoder or corrector circuitry can at most corrupt a single codeword bit (i.e., no single error can propagate to multiple codeword bits);
2) There is a fault secure detector that can detect any combination of errors in the received codeword along with errors in the detector circuit. This fault-secure detector can verify the correctness of the encoder and corrector operation. The first property is easily satisfied by preventing logic sharing between the circuits producing each codeword bit or information bit in the encoder and the corrector respectively. In the next sections, we define the requirements for a code to satisfy the second property.

A. *Euclidean geometry codes*

This section reviews the construction of Euclidean Geometry codes based on the lines and points of the corresponding finite geometries. Euclidean Geometry codes are also called EG-LDPC codes based on the fact that they are low-density parity-check (LDPC) codes. LDPC codes have a limited number of 1's in each row and column of the matrix. This limit guarantees limited complexity in their associated detectors and correctors making them fast and light weight.

Let EG be a Euclidean Geometry with n points and J lines.EG is a finite geometry that is shown to have the following fundamental structural properties:

1) Every line consists of r points;
2) Any two points are connected by exactly one line;
3) Every point is intersected by g lines;
4) Two lines intersect in exactly one point or they are parallel; i.e., they do not intersect.

Let H be a J x n binary matrix, whose rows and columns corresponds to lines and points in an EG Euclidean geometry respectively, where $h_{i,j} = 1$ if and only if the i^{th} line of EG contains the j^{th} point of EG, and $h_{i,j} = 0$ otherwise. A row in H displays the points on a specific line of EG and has weight r. A column in H displays the lines that intersect at a specific point in EG and has weight g . The rows of H are called the incidence vectors of the lines in EG, and the columns of H are called the intersecting vectors of the points in EG. Therefore, H is the incidence matrix of the lines in EG over the points in EG .It is shown that H is a LDPC matrix, and therefore the code is an LDPC code.

A special subclass of EG-LDPC codes, type-I 2-D EG-LDPC, is considered here. The theory behind EG-LDPC codes are explained in [3].It is shown in that type-I 2-D EG-LDPC has the following parameters for any positive integer $t > 2$:

- Information bits, $k=2^{2t} - 3^t$;
- Length, $n =2^{2t} - 1$;
- Minimum distance, $d\ min =2^t+1$;
- Dimensions of the parity-check matrix, n x n;
- Row weight of the parity-check matrix, $p =2^t$;
- Column weight of the parity-check matrix, $y =2^t$

B. *FSD-ECC definition*

The restricted ECC definition which guarantees a fault-secure detector capable ECC (FSD-ECC) is as follows: Let C be an ECC with minimum distance d. C is FSD-ECC if it can detect any combination of overall $d-1$ or fewer errors in the received codeword and in the detector circuitry.

Let C be an ECC, with minimum distance d. C is FSD-ECC iff any error vector of weight $e \leq d-1$, has syndrome vector of weight at least $d-e$.

The following proof depends on the fact that any single error in the detector circuitry can corrupt at most one output (one syndrome bit). This can be easily satisfied for any type of circuitry by implementing the circuit in such a way that no logic element is shared

among multiple output bits, therefore any single error in the circuit corrupt at most one output (one syndrome bit).

Proof: The core of a detector circuitry is a multiplier that implements the vector-matrix multiply of the received vector and the parity-check matrix to generate the syndrome vector.

Now if e errors strike the received code-vector the syndrome weight of the error pattern is at least d − e from the assumption. Furthermore, the maximum number of tolerable errors in the whole system is d − 1 and e errors already exist in the encoded vector, therefore the maximum number of errors that can strike in the detector circuitry is d − 1 − e. From the above note, this many errors can corrupt at most d − 1 − e syndrome bit, which in worst case leaves at least one non-zero syndrome bit and therefore detects the errors.

The difference between FSD-ECC and normal ECC is the demand on syndrome weight: i.e., a normal ECC demands non-zero syndrome weight while FSD-ECC demands ≥ d − e.

C. FSD-ECC Proof for EG-LDPC codes

Any ECC of type-I 2-dimensional EG-LDPC is FSD-ECC.

Proof: Let C be an EG-LDPC code with column weight g and minimum distance d. We have to show that any error vector of weight e ≤ d−1 corrupting the received encoded vector has syndrome vector of weight at least d − e. A specific bit in the syndrome vector is 1 if and only if the parity-check sum corresponding to this syndrome vector has an odd number of error bits present in it. Looking from the Euclidean geometry perspective, each error bit corresponds to a point in the geometry and each bit in the syndrome vector corresponds to a line. Now we are interested in obtaining a lower bound on the number of lines that pass through an odd number of error points. We further lower bound this quantity by the number of lines that pass through exactly one of the error points. Based on the definition of the Euclidean geometry, g lines pass through each point; so e error points potentially impact ge lines. Also at most one line connects two points. Therefore, looking at the e error points, there are at most $\binom{e}{2}$ lines between any two error points. Hence the number of lines passing through a collection of these e points is lower bounded by e $-\binom{e}{2}$. Out of this number, at most $\binom{e}{2}$ lines connect two or more points of the error points. Summarizing all this, the number of lines passing through exactly one of the error points is at least ge $- 2\binom{e}{2}$. Note from the code properties of EG-LDPC codes that d = g+ 1, we can derive the following inequality: ge $- 2\binom{e}{2}$= e(g+ 1 − e) = e(g − e) ≥ d − e Which prove that the weight of the syndrome vector is at most d − e which is the required condition of Theorem (I). Therefore EG- LDPC are FSD-ECC.

III. FAULT-TOLERANT MEMORY SYSTEM

The architecture of the proposed reliable, Fault-tolerant memory system is shown in the figure 1 and is described in the following.

Here, we outline our memory system design that can tolerate errors in any part of the system, including the storage unit and encoder and corrector circuits using the fault-secure detector. The information bits are fed into the encoder to encode the information vector, and the fault secure detector of the encoder verifies the validity of the encoded vector. If the detector detects any error, the encoding operation must be redone to generate the correct codeword. The codeword is then stored in the memory. During memory access operation, the stored code words will be accessed from the memory unit. Code words are susceptible to transient faults while they are stored in the memory; therefore a corrector unit is designed to correct potential errors in the retrieved code words. In our design all the memory words pass through the corrector and any potential error in the memory words will be corrected. Similar to the encoder unit, a fault-secure detector monitors the operation of the corrector unit.

All the units shown in Fig. 1 are implemented in fault-prone circuitry; the only component which must be implemented in reliable circuitry are two OR gates that accumulate the syndrome bits for the detectors.

Data bits stay in memory for a number of cycles and during this period, each memory bit can be upset by a transient fault with certain probability. Therefore, transient errors accumulate in the memory words over time. In order to avoid accumulation of too many errors in any memory word that surpasses the code correction capability, the system must perform memory scrubbing.

Memory scrubbing is the process of periodically reading memory words from the memory, correcting any potential errors, and writing them back into the memory. To perform the periodic scrubbing operation, the normal memory access operation is stopped and the memory performs the scrub operation.

IV. ERROR DETECTION AND CORRECTION PROCESS

The error detection and correction process involves the design of the encoder, fault-secure detector and corrector.

A. Design of the encoder:

An n-bit codeword C, which encodes a k-bit information vector I is generated by multiplying the k-bit information vector with a k x n bit generator matrix G. The code rate is defined as the fraction k/n of k source symbols and n encoded symbols.

$$\text{i.e., } C = I \,.G \tag{4.1}$$

EG-LDPC codes are not systematic and the information bits must be decoded from the encoded vector, which is not desirable for our fault-tolerant approach due to the further complication and delay that it adds to the operation. However, these codes are cyclic codes. A code is a systematic code if any codeword consists of the original k-bit information vector followed by (n – k) parity-bits. The advantage of using systematic codes is that there is no need for a decoder circuitry to extract the information bits. The information bits are simply available in the first *k* bits of any encoded vector. With this definition, the generator matrix of a systematic code must have the following structure.

$$G = [I: X] \tag{4.2}$$

Where I is a k × k identity matrix and

X is a k× (n−k) matrix that generates the parity-bits.

EG-LDPC codes have the following parameters for any positive integer t > 2: where t is the number of errors that the code can correct.

These parameters are explained in the below section in the design of the detector.

For example consider t =2,

- Information bits, $k=2^{2t} – 3^t = 7$;
- Length, $n =2^{2t} – 1 = 15$;
- Minimum distance, $d\ min =2^t +1 = 5$;
- Dimensions of the parity-check matrix, n x n = 15x15;
- Row weight of the parity-check matrix, $p =2^t = 4$;
- Column weight of the parity-check matrix, $y =2^t = 4$; Code rate = k/n = 7/15

Then generator matrix formed as follows.

$$G = [I: X] \tag{4.3}$$

Where I is a k × k = 7 x 7 identity matrix and

X is a k× (n−k) = 7 x 8 matrix that generates the parity-bits

The generator polynomial of (15,7,5) EG-LDPC code is given by $1 + x^4 + x^6 + x^7 + x^8$

This generator polynomial will result in the generator matrix in cyclic format shown in the figure below.

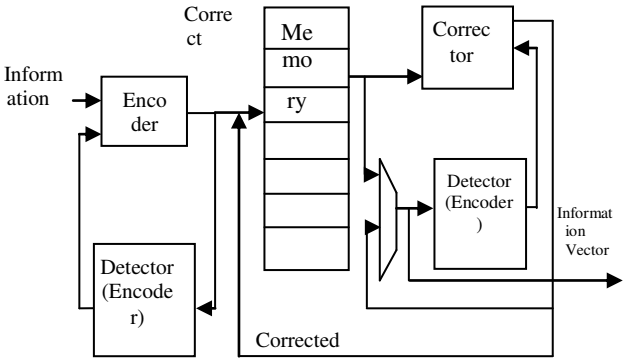


Fig. 1 : The overview of the proposed fault tolerant memory architecture

	c_0	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	c_{14}
i_0	1	0	0	0	1	0	1	1	1	0	0	0	0	0	0
i_1	0	1	0	0	0	1	0	1	1	1	0	0	0	0	0
i_2	0	0	1	0	0	0	1	0	1	1	1	0	0	0	0
i_3	0	0	0	1	0	0	0	1	0	1	1	1	0	0	0
i_4	0	0	0	0	1	0	0	0	1	0	1	1	1	0	0
i_5	0	0	0	0	0	1	0	0	0	1	0	1	1	1	0
i_6	0	0	0	0	0	0	1	0	0	0	1	0	1	1	1

Fig. 2 : The generator matrix of (15, 7, 5) EG-LDPC code in cyclic format.

We used the procedure presented in [2] and [3] to convert the cyclic generator matrices to systematic generator matrices for all the EG-LDPC codes under consideration. The encoded vector consists of information bits followed by parity bits, where each parity bit is simply an inner product of information vector and a column of parity-check matrix. Fig.3 shows the encoder circuit to compute the parity bits of the (15, 7, 5) EG-LDPC code. In this figure the information vector will be copied to bits of the encoded vector and the rest of encoded vector, the parity bits, are linear sums (XOR) of the information bits

The core of the detector operation is to generate the syndrome vector, which is basically implementing the following vector-matrix multiplication on the received encoded vector and parity-check matrix :

$$\text{i.e., } S = C\,H^T \tag{4.4}$$

Therefore each bit of the syndrome vector is the product of received encoded vector bit with one row of the parity-check matrix. This product is a linear binary sum over digits of received encoded vector, where the corresponding digit in the matrix row is 1. This binary sum is implemented with an XOR gate. Fig. 4 shows the detector circuit for the (15, 7, 5) EG-LDPC code since the row weight of the parity-check matrix is

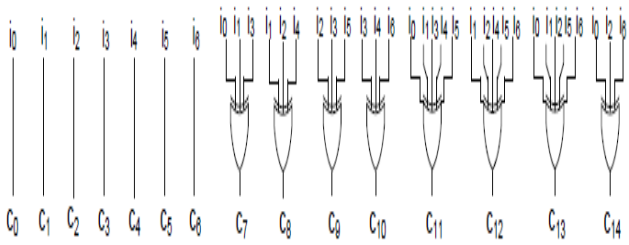


Fig. 3 : Structure of the Encoder circuit

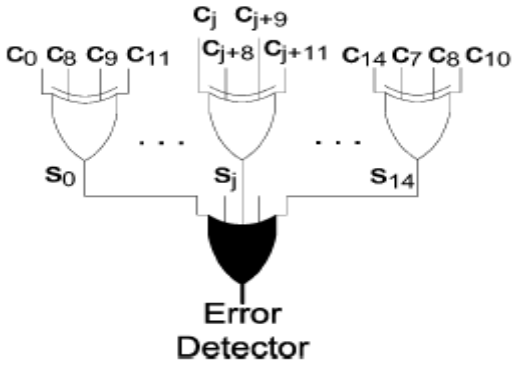


Fig. 4 : Fault secure detector of the (15,7,5) EG-LDPC code

ρ, to generate one digit of the syndrome vector, we need an ρ input xor gate, or (ρ-1) 2-inputs xor gates. For the whole detector it takes n(ρ-1) 2-input xor gates. Here, each syndrome bit is implemented with a separate xor gate satisfies the assumption of Theorem 1 of no logic sharing in detector circuit implementation.

An error is detected if any of the syndrome bits has a non-zero value. The final error detection signal is implemented by an OR function of all the syndrome bits. The output of this n-input OR gate is the error detector signal. In order to avoid single point of failure, we must implement the OR gate with a reliable substrate (example: in a system with sub-lithographic nano-wire substrate.).

B. Design of the decoder

One-step majority-logic correction is a fast and relatively compact error-correcting technique [4]. There is a limited class of ECCs that are one-step-majority correctable which include type-I two-dimensional EG-LDPC. In this section, we present a brief review of this correcting technique. Then we show the one-step majority-logic corrector for EG-LDPC.

This method consists of two parts: 1) generating a specific set of linear sums of the received vector bits and 2) finding the majority value of the computed linear sums. The majority value indicates the correctness of the code-bit under consideration; if the majority value is 1, the bit is inverted, otherwise it is kept unchanged.

The theory behind the one-step majority corrector and the proof that EG-LDPC codes have this property are available in [4]. Here we overview the structure of such correctors for EG-LDPC codes.

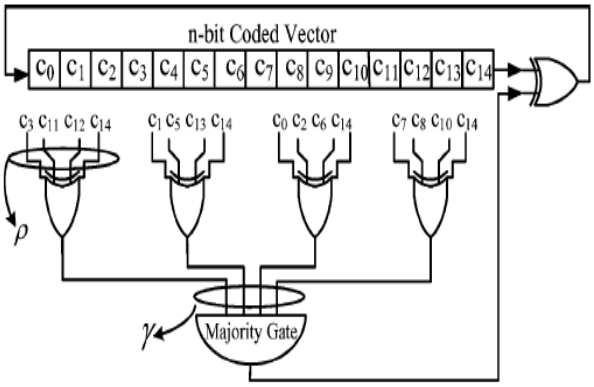


Fig. 5: Structure of the serial one step majority corrector

A linear sum of the received encoded vector bits can be formed by computing the inner product of the received vector and a row of a parity-check matrix. This sum is called *Parity-Check* sum. The core of the one-step majority-logic corrector is generating parity-check sums from the appropriate rows of the parity-check matrix. The one-step majority logic error correction is summarized in the following procedure.

These steps correct a potential error in one code bit.

1) Generate g parity-check sums by computing the inner product of the received vector and the appropriate rows of parity-check matrix.

2) The g check sums are fed into a majority gate. The output of the majority gate indicates the correctness of the code bit under consideration; if the majority value is 1, the bit is inverted, otherwise it is kept unchanged.

The circuit implementing a serial one-step majority logic corrector for (15, 7, 5) EG-LDPC code is shown in Fig 5. This circuit generates g = 4 parity-check sums with g = 4 XOR gates and then computes the majority value of the parity-check sums. Since each parity-check sum is computed using a row of the parity check matrix and the row density of EG-LDPC codes are ρ = 4, each XOR gate that computes the linear sum has ρ = 4 inputs. The single XOR gate on the right of the above Fig corrects the code bit c_{n-1} using the output of the majority gate. Once the code bit c_{n-1} is corrected the codeword is cyclic shifted and code bit c_{n-2} is placed at position c_{n-1}

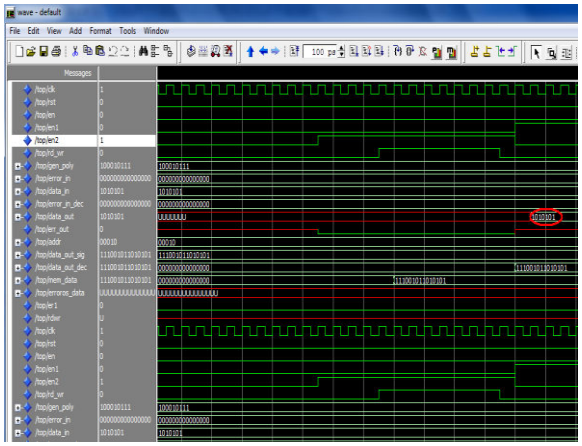
and will be corrected. The whole codeword can be corrected in *n* rounds.

V. RESULTS

The proposed system can be analysed by running below simulations.

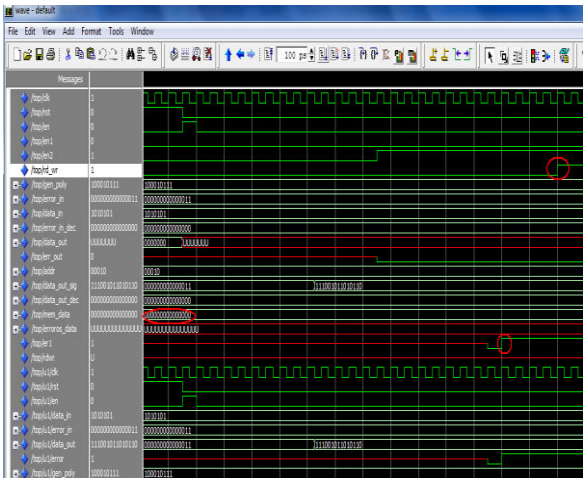
- The system with no errors anywhere.
- The system with errors at the encoder or input.
- The system with errors in memory, decoder or output

a. The system with no errors anywhere.



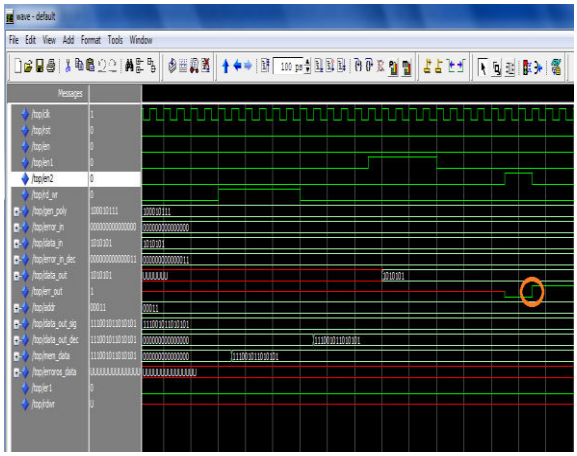
Since there are no errors within the entire system, we get correct output data**.**

b. The system with errors at the encoder or input.



Due to the errors introduced during the encoding operation, the encoded data will not be written in to the memory even though, rd_wr signal is high. Therefore the encoder must repeat the operation for the correct encoded data.

c. The system with errors in memory, decoder or output.



The errors introduced in the memory or during the operation of the corrector are detected by the detector at the corrector.

VI. CONCLUSIONS

In this project work, the design and implementation of a fault-tolerant memory system is carried out, such that the detection and correction of errors is done both in the storage unit and in the supporting circuitry such as encoder and decoder. The concept of Fault–Secure Detector capable ECC (FSD-ECC) is defined, which allows the implementation of memory systems with protected supporting circuitry. The FSD-ECC property of Euclidean geometry codes are proved and computed the reliability of the complete system. The design of such a memory system is desirable for high fault-tolerant applications such as servers, as well as deep-space applications due to increased radiations. It is also useful in the data storage devices such as compact disc, hard disc, DVDs where the stored data is replayed in later time. The functional verification of the proposed system is carried out by running the simulations in Modelsim. The simulation results are attached for verification. And also the implementation is done on spartan3 FPGA.

VII. FUTURE SCOPE

The future enhancement and scope involves identifying the area overhead of the proposed system with the unprotected system and need to compare the failure rates of the memory system and supporting logic.

REFERENCES

[1] S. Hareland, J. Maiz, M. Alavi, K. Mistry, S. Walsta, and C. Dai, “Impact of CMOS process scaling and SOI on the soft error rates of logic

processes," in Proc. Symp. VLSI, 2001, pp. 73–74.

[2] Y. Kou, S. Lin, and M. P. C. Fossorier, "Low-density parity-check codes based on finite geometries: A rediscovery and new results," IEEE Trans. Inf. Theory, vol. 47, no. 7, pp. 2711–2736, Jul. 2001.

[3] S. Lin and D. J. Costello, Error Control Coding, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 2004.

[4] R. J. McEliece, The Theory of Information and Coding. Cambridge, U.K.: Cambridge University Press, 2002.

[5] H. Naeimi and A. DeHon, "Fault secure encoder and decoder for memory applications," in Proc. IEEE Int. Symp. Defect Fault Tolerance VLSI Syst., Sep. 2007, pp. 409–417.

[6] H. Naeimi and A. DeHon, "Fault-tolerant nano-memory with fault secure encoder and decoder," presented at the Int. Conf. Nano-Netw., Catania, Sicily, Italy, Sep. 2007.

[7] S. J. Piestrak, A. Dandache, and F. Monteiro, "Designing fault-secure parallel encoders for systematic linear error correcting codes," IEEE Trans. Reliab., vol. 52, no. 4, pp. 492–500, Jul. 2003.

[8] A. Saleh, J. Serrano, and J. Patel, "Reliability of scrubbing recovery techniques for memory systems," IEEE Trans. Reliab., vol. 39, no. 1, pp. 114–122, Jan. 1990.

[9] Synthesis and Simulation Design Guide, Xilinx.

[10] Simulation Design Guide, Mentor Graphics.

◈◈◈

Gnome Sorted Hessian Percolation Model For Automated Visual Quality Inspection of Manufactured Metallic Products

Sarin CR ME (Mechatronics)
Karppagam College of Engineering
E-mail : crsarin@gmail.com

I. INTRODUCTION

Quality assurance is the systematic monitoring and evaluation of the various aspects of a product to maximize the probability that minimum standards of quality are being attained by the production process. Deviations from the normal quality that impair the operating characteristics of a metal or product and lead to a reduction in grade or to rejection of products should be considered as defects [1]. Imperfections in the manufactured products may build up during smelting of the metal and production of castings; during pressure treatment; as a result of thermal, chemicothermal, electrochemical, and mechanical treatment and in the process of joining metals like during welding, soldering, riveting, and so on. Metal defects may be local, distributed in limited zones, or distributed throughout the entire volume or surface of a product [2]. They may turn out negatively on various properties such as electrical conductivity, magnetic permeability, strength, density, and plasticity etc.

Due the high expectations of both primary manufacturers and end consumers, defects cannot be tolerated even in million piece quantities with considerable interest in optimum use of materials and cost reduction. The requirement for defect free manufacturing has driven suppliers to seek out cost effective methods of 100% quality. With introduction of Quality and Management Systems to industries has led to major improvements in the reliability of plant as well as their availability, but mainly focused to be as good as the data used as input for these systems [3]. To achieve zero defects ("Zero PPM") output cost-effectively, manufacturers are making the commitment to move to online, automated Non Destructive Testing (NDT) methods. This type of online inspection requires accuracy, reliability, and high throughput. The quality monitored by NDT methods such as chemical, spectral, structural, and metallographic analysis is difficult to be performed in real-time operation [4]. More reliable, thorough monitoring is accomplished through the advanced image processing techniques in NDT by the structural analysis of products. Gnome Sorted Hessian Percolation Model based automated visual quality inspection and NDT is a very efficient method of structural defect discrimination and lends itself very well to 98% inspection of industrial imperfections.

Automatic defect recognition (ADR) in the case of NDT, especially by Real Time Radioscopy is rapidly become the accepted way for controlling the quality through visual or computer-aided analysis of images automatically [5]. The purpose of this non-destructive testing method is to identify microscopic casting defects and cracks automatically, which may be even internal fault in nature and measure them by intelligent object detection and feature extraction tools. The AVI is a quality control task to determine automatically whether a casting complies with a given set of product and product safety specifications [6].

There has been a remarkable development in research in the field of radiography in the last few decades, resulting in many new technologies like Real Time Radioscopy, Digital Radiography, and Computed Radiography etc. These inventions have significantly enhanced the quality and productivity of Radiographic testing through reduction in cycle time by elimination of chemical film processing, and image processing applications [7].

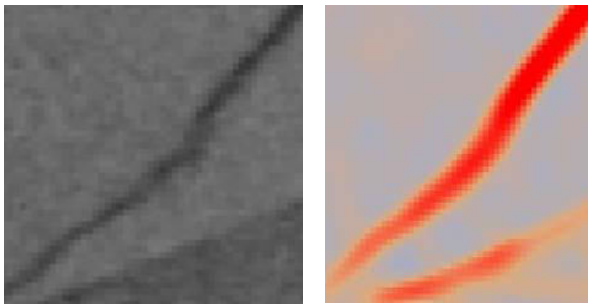


Fig. 1: Voids in radioscopic images of metal wheels and enhanced image of defect.

II. COMPUTER AIDED RADIOGRAPHIC TESTING

The principle aspects of an automated inspection unit are shown in Fig. 1.2. Typically, it comprises the following five steps [8]:

1. A manipulator for handling the test piece,
2. An source, which irradiates the test piece with a conical beam to generate an image of the test piece,
3. An image intensifier which transforms the invisible image into a visible one,
4. A CCD camera which records the visible image and
5. A computer to process the digital image processing of the image and then classifies the test piece accepting or rejecting it.

Programmable controller (PLC) may also control the manipulator for positioning the test piece in the desired inspection position, although this task is normally performed by a computer.

Detectors made of compound semiconductors such as CdTe and CdZnTe have shown outstanding performance for X and gamma ray spectrometry when operating at room temperature [9].

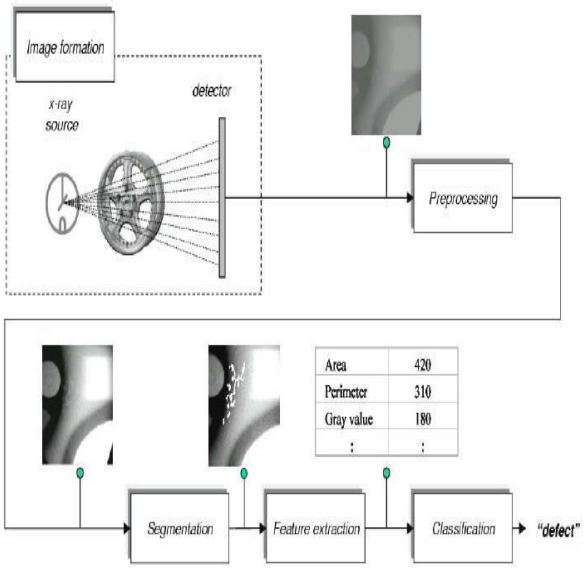


Fig. 2: Phases of pattern recognition in automated flaw detection.

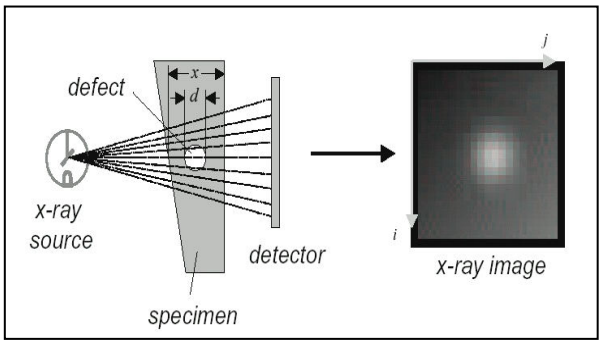
This paper proposes an idea of using intelligent object detection and feature extraction in image processing as a tool in the automated visual inspection and NDT of finished products. The paper introduces the Gnome Sorted Hessian Percolation Model based automated visual quality inspection and NDT employed when inspecting metal products.

III. DIGITAL IMAGE PROCESSING IN TESTING

The aim of the Digital image processing system is to recognize the fault and its dimensions and features using artificial intelligent technique. The AI classifies the image obtained from the ITC is processed and pattern reorganization algorithms are executed to get to the classification of image elements: 'Area with defects' and 'Background' [10-12].

3.1 Image Formation

The more common way of acquisition is through scanners, which works with light transmission – usually called transparency adapters. The material under test is exposed to radio waves and ITC detectors made of compound semiconductors such as CdTe and CdZnTe senses the radiation intensity attenuated by the material. Another method also used is image acquisition by camera CCD (Charge Couple Device). A defect in the material modifies the expected radiation received by the sensor. The intensity distribution of is characterized by structural distribution of the product and defects such as voids, cracks or bubbles, show up as bright features which patterns a varying intensity provision image due to low attenuation.



The real-time acquisition of image in matrix representation and the size of the matrix correspond to the resolution of the image. Proposed system resolution is 286 × 384 pixels, each associated a value, usually for gray scale images it is between 0 and 255 for a scale of 28 = 256 gray levels. Here, '0' represents 100% black and a value of '255' corresponds to 100% white, matrix x be the digitized image, then the element x(i,j) denotes the gray value. The eye is only capable of resolving around 40 gray levels, however for the detection of ADC system, 216 = 65,536 gray levels are used, which allows one to evaluate both very dark and very bright regions in the same image increases system performance.

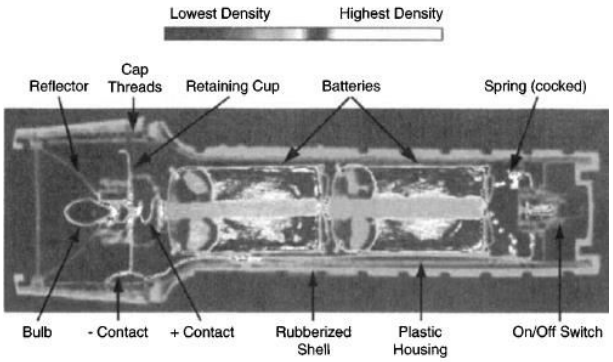


Fig. 3: Differential absorption in a specimen.

3.2 Preprocessing

The image taken is preprocessed to improve the quality of the image before it is analyzed removes noise, enhanced the contrast, corrected the shading effect and restore blur deformation of images, seeking mainly the attenuation by elimination of noise and contrast improvement.

3.2.1 Noise Removal

Noise follows Poisson law directly which can prove a significant source of image degradation and taken into consideration account during processing and analysis. The standard deviation of this distribution is equal to the square root of the mean. This means that the photon noise amplitude is signal-dependent. Integration removes image noise stationary images modeled using the stationary component and the noise component With a noise component of zero mean, average of the n images the stationary component is unchanged, while the noise pattern decreases by increasing n improves the signal-to-noise ratio by a factor. The larger the number n, the better the improvement, usually there added ($10 \leq n \leq 16$).

3.2.2 Contrast Enhancement

Contrast encasement will amplify the differences in the gray levels of the image whose gray values lay in a relatively narrow range of the gray scale. The function summarizes the gray level information of an image as histogram h(x) where x is a gray level and h(x) denotes the number of pixels in the image that have a gray level equal a definite value. Simple contrast enhancement can be achieved if we use a linear transformation which sets the minimal and maximal gray values of the image to the minimal and maximal gray value of the gray level scale respectively, the gray levels expand from '0' to '255'.The nonlinear transformation is usually performed with a γ-correction so that $\gamma > 1$ the mapping is weighted toward darker output values, and $\gamma < 1$ the mapping is weighted toward brighter output values.

3.2.3 Shading Correction

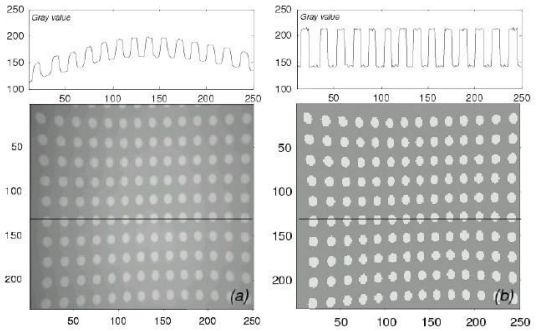
A decrease in the angular intensity in the projection causes low spatial frequency variations in Since the plate is of a constant thickness, we would expect to see a constant gray value for the metal part and another constant gray value for the holes, can be overcome by using linear shading correction[13].

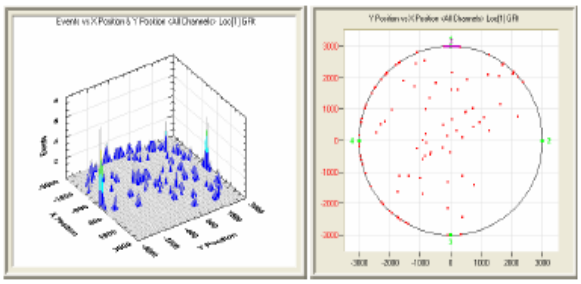
Input image x(i,j) will be linear transformed according to

$$y(i,j) = a(i,j)\, x(i,j) + b(i,j)$$

The coefficients a(i,j) and b(i,j) are estimated by analyzing two real images r1(i,j) and r2(i,j) and the corresponding ideal images i1(i,j) and i2(i,j)

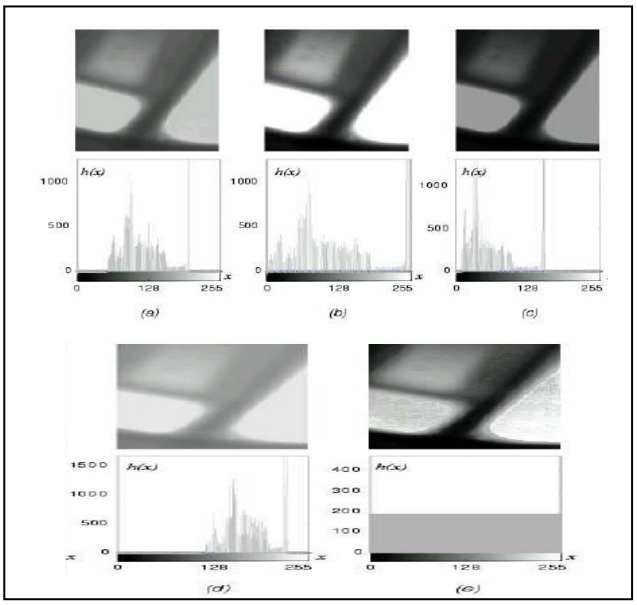
$$ik(i,j) = a(i,j)\, rk(i,j) + b(i,j) \text{ for } k = 1,2,3\ldots\ldots.$$





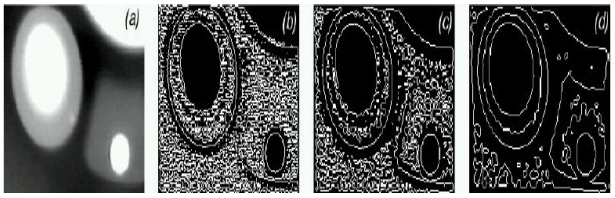
3.2.4 Restoration of blur caused by motion

Recovering detail in severely blurred images for known a-priori images exist as an analytical model, or as a-priori information in conjunction with knowledge (or assumptions) of the physical system that provided the imaging to estimate the best source image, where there given the blurred example and some a-priori knowledge. The blur caused by uniform linear motion is removed by assuming that the linear motion corresponds to an integer number of pixels and is horizontally (or vertically) aligned with sampling raster. In these examples, the details of the metal castings are not discernable in the degraded images, but are recovered in the restored image.



3.3 Segmentation

Image segmentation is defined as the process of subdividing an image into disjointed regions. In image processing for detecting faults in castings, such regions correspond to potential defects and the background (or regular structures). While there are many methods for segmenting images, two approaches for segmenting potential defects in images are used widely within the nondestructive testing community. The first technique is based on median filtering while the second is a region-oriented method [12].

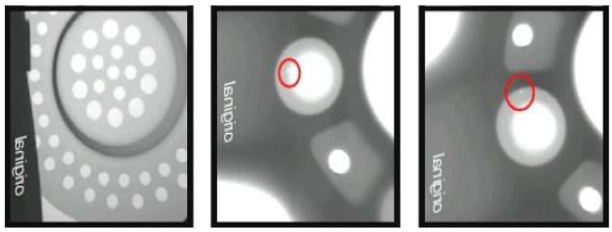


3.4 Edge Detection and Region Finding

This approach attempts to detect the potential defects in an image in two steps: edge detection and region finding .In edge detection, the edges of the image are detected, correspond to pixels of the image in which the gray level changes significantly over a short distance. The edges are normally detected using gradient operators. In the second step, the regions demarcated by the edges are extracted. The key idea of region finding two step based approach is that regions demarcated by the edges are extracted by the existing defects with present significant gray level changes compared to their surroundings. A Laplacian of Gaussian (LoG) kernel and a zero crossing algorithm can be used to detect the edges of the images. The LoG-operator involves a Gaussian low pass filter, which is good for the pre-smoothing of the noisy images. The LoG-kernel depends on parameter σ, which defines the width of the Gaussian function and, thus, the amount of smoothing and the edges detected using the LoG-kernel we calculate an image in which the edges of the original image are located by their zero crossing. The detected edges correspond to the maximal (or minimal) values of the gradient image. The binary edge image obtained should reproduce real flaws' closed and connected contours that demarcate regions

3.5 Feature extraction and selection

An analysis of the segmented regions, however, can improve the effectiveness of fault detection significantly by measuring certain characteristics of the segmented regions (feature extraction) can help us to distinguish the defects, although some of the extracted features are either irrelevant or are not correlated. Therefore, a feature selection must be performed. Depending on the values returned for the selected features, there classified each segmented region in one of the following two classes: 'regular structure' or 'defect'.

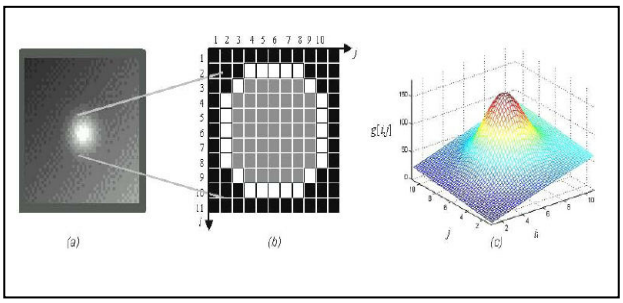


3.5.1 Feature extraction

Features that are normally used in the classification of potential defects, usually divided into two groups: geometric and gray value features. Geometric features provide information about the size and the shape of the segmented potential flaw. The extracted geometric features can be: area, perimeter, height, width, roundness, Hu invariant moments, Flusser and Suk invariant moments, Fourier descriptors, semi-minor and semi-major axis of ellipse fitted to the contour of the potential flaw, and Danielsson shape factor.

The gray value features provide information on the brightness of the segmented potential flaw where the extracted features are: mean gray value, mean gradient in the boundary, mean second derivate in the region, radiographic contrasts, contrasts based on crossing line profiles, invariant moments with gray value information, local variance, mean and range of the Haralick textural features (angular second moment, contrast, correlation, sum of squares, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, information measures of correlation and maximal correlation coefficient) based

on the co-occurrence matrix in four different directions taken neighboring pixels separated by several distances, and components of the discrete Fourier transform, the Karhunen Loève transform and the discrete cosine transform taken from a window including potential flaw and neighborhood [13].



3.5.2 Gnome Sorted Hessian Percolation Algorithm

Several techniques for crack detection have been developed recently. The most obvious solution is the use of standard intelligent image processing methods or combinations of it. Gnome sort proposed by Hamid Sarbazi-Azad [14] in 2000 then later on described by Dick Grune and named "Gnome sort" [15], is a sorting algorithm where moving an element to its proper place is accomplished by a series of swaps. The running time is O(n²), but tends towards O(n) if the list is initially almost sorted with an he average runtime is O(n2).The algorithm always finds the first place where two adjacent pixel elements are in the wrong order, and swaps them. It takes advantage of the fact that performing a swap can introduce a new out-of-order adjacent pair only right before or after the two swapped elements. It does not assume that elements forward of the current position are sorted, so it only needs to check the position directly before the swapped elements.

Hessian-Driven Percolation is an advanced method suitable for crack detection which gives a smoothed image to obtain uniform brightness, followed by removing isolated points to remove noise and morphological operations with fast operation [16-18]. The percolation algorithm described in works can be used for even complex images which are based on the physical model of liquid permeation, are started from each pixel. Depending on the shape of the percolated region, the pixel is considered as a crack pixel or not. Basically the simple percolation process consists of the described syntax:

1. Initialize the genome operator *i*=1 and the latest iterative variable *k* to 0
2. If the genome operator value is less then the maximum gray level value n, following conditions are executed.
3. The starting pixel P_s is added to the set of percolated pixels *P*
4. The threshold *t* is set to the value

$$t = max\left(\int_{p\in Dp}^{\infty} max\ I(p), t\right) + w,$$

Where w is a parameter to accelerate the percolation

5. Each pixel p neighboring P is added P to if I(P)≤ t
6. The circularity *Fc* is defined by

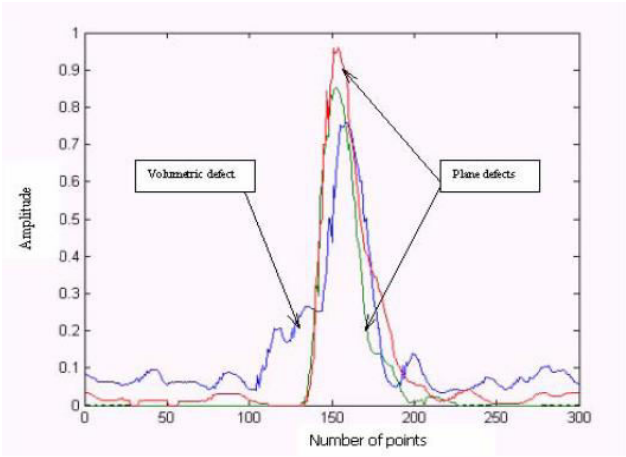
$$Fc = {4 \times |P|}/{\pi(diam\ P)^2}$$

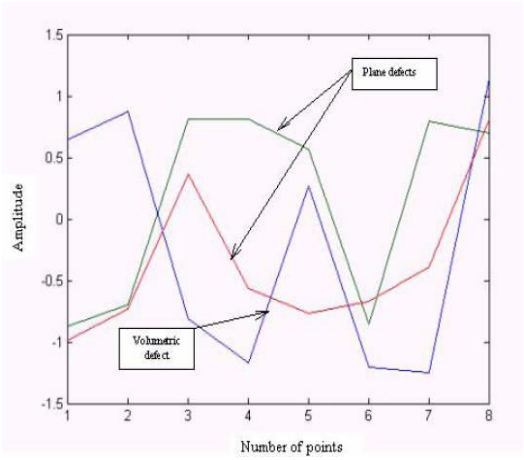
7. Genome operator is redefined by the value of k and k is assigned to zero If the P exceed the preceder, P> P(n-1)
8. Pixel values are interchanged and k redefined by the value of genome operator and *i* is assigned to zero if, P< P (n-1).
9. Genome operator is increased in each step by one for all values n less than *i* else decreased by one.
10. Iteration on k is implemented until p reaches boundary of M×M window with centre P if the circularity *Fc* of *P is* close to 0, p is considered as a crack pixel.

The percolation algorithm described in the previous section cannot be used immediately for the detection of cracks in complex images. While steps 1–4 can be directly applied to complex images, the circularity *Fc* computed in step 5 has no obvious generalization to complex images. An extension of the percolation algorithm to complex images with improved computation time is proposed in this section. To overcome the problems mentioned above, we employ the sheet filter. Let H be the result obtained from by S choosing an appropriate threshold.

$$Fc \text{ is determined by} = {|P \cap H|}/{P}$$

The computation of the circularity of the percolated region is replaced by the computation of we check if the percentage of pixels in which *H* is contained in is close to 1.Furthermore we use *H* to improve the speed of the calculation. This is achieved by starting the percolation process only at pixels contained in $(|P \cap H|) / P$. Note that if we considered the starting pixel of the percolation process as a crack pixel only if is close to 1, then the result of the percolation algorithm would be a subset of. Thus it would be





3.5.3 Feature selection

In feature selection we have to decide just which features of the regions are relevant to the classification so that n extracted features are arranged in an n -vector that can be viewed as a point in an n-dimensional space. In addition, each feature can be considered as a random variable with no samples. Each variable is normalized in order to obtain a zero mean and a standard deviation equal to one.

The key idea of the feature selection is to select a subset of m features (m < n) that leads to the smallest classification error with selected m features are arranged in a new m -vector. The selection of the features can be done using Sequential Forward Selection which selects the best single feature and then adds one feature at a time that, in combination with the selected features, maximizes classification performance so that iteration is stopped once no considerable improvement in the performance is achieved on adding a new feature. By evaluating selection performance we ensure:

a. A small intraclass variation and

b. A large interclass variation in the space of the selected features.

For the first condition the intraclass-covariance matrix is used, and for the second the covariance matrix of each class. The best features that separate the classes 'defects' and 'regular structures' are related to the contrast

IV. CLASSIFICATION

Once the proper features are selected, a classifier can be designed which typically assigns a feature vector to one of the two classes: 'regular structure' or 'defect', that are assigned ‘0’ and ‘1’, respectively. In statistical pattern recognition, classification is performed using the concept of similarity: patterns that are similar are assigned to the same class though this approach is very simple, a good metric defining the similarity must be established. Using a representative sample we can make a supervised classification finding a discriminant function that provides us information on how similar a feature vector is to the feature vector of a class [19].

V. CONCLUSION

The importance of quality control and nondestructive inspection are known-well for industrial applications because of safety, very high cost and complexity of manufacturing technology as well as time-cost. One of the biggest difficulties in NDT of these structures are time-cost and high quality control requirements which are achieved in an improved by using the Hessian-Driven Percolation based automated visual quality inspection and NDT technique. The developed high-efficient automated scanning imaging technique realizes a fast NDT for the metallic structures which gives an advanced radiographic scan imaging technique by employing a novel radiographic matrix up to 20 transducer (SITC) elements, and is developed to provide a multiple channels scanning for even large-scale complex structures. Each transducer (SITC) element in the matrix can independently and self-adapted to follow the structures to be tested during automated scanning with the help of preprogrammed reliable control system. The scanning area is extended up to 6000mm in width, and unlimited in length. The practical and industrial applications have demonstrated the powerful ability and flexibility as well as high-efficiency in the NDT of large-scale metallic structures. The inspection efficiency is increased up to 15-20 times compare conventional radiographic scanning technique.

REFERENCE

[1] Juran, J. M., Godfrey, A. B., Juran’s Quality Handbook, 7th ed. McGraw Hill, New York, NY (2006)

[2] Kisil’ I.S., Karpash M.O., Vashchyshak I.R. Device for Evaluation of Mechanical Properties

of Steel FMH-1 - Methods and Devices of Quality Control, 2005, No.14, 77-80.

[3] Charola, A.E., E. De Witte, and M. Laurenzi Tabasso. Establishing International Standards for the Quality Control of Conservation Materials and for Qualifying Practioners Applying Them. In Saving Our Architectural Heritage. 1996. Berlin: Wiley; p. 245–254.

[4] Hands, G., Nath, R. H., "Nondestructive Testing of Metal Castings," Materials Evaluation, vol 62, No 2, pp 111–115 (2004).

[5] Artificial Intelligence: A Modern Approach, Stuart J. Russell and Peter Norvig, Prentice Hall, Englewood Cliffs, ISBN 0-13-080302-2.

[6] Dario Floreano, Toshifumi Kato, Davide Marocco, and Eric Sauser. Automaitic vision and feature selection. Biological Cybernetics, 90:218–228, 2004

[7] S.W. Lawson. Automatic defect detection in industrial radioscopic and ultrasonic images. PhD. thesis, University of Surrey, 1996.

[8] S. Edelman; Representation and Recognition in Vision, MIT Press, 1999, ISBN 0-262-05057-9.

[9] Ultrasonic and Advanced Methods for Nondestructive Testing and Material Characterization, World Scientific Publishing Co. Pte. Ltd,2005.

[10] Digital Image Processing,Rafael C. González, Richard Eugene Woods,Prentice Hall, 2008 - Computers - 954 pages

[11] Computer vision, graphics and image processing: fourth Indian conference, ICVGIP 2004, Kolkata, December 16-18, 2004. Proceedings Chanda, B., ed., Chandran, S., ed. & Davis, L., ed. 006.37 CHA 019131

[12] F. Brandt, 2000. The use of X-ray inspection techniques to improve quality and reduce costs. The e-Journal of Nondestructive Testing & Ultrasonics, 5(5).

[13] [A. Besbes 2009] N. Paragios A. Besbes N. Komodakis and G. Langs. Discrete MRFs for Knowledge-based Segmentation. IEEE Conference in Computer Vision and Pattern Recognition, 2009. 103

[14] [Amberg 2007] B. Amberg, S. Romdhani and T. Vetter. Optimal Step Nonrigid ICP Algorithms for Surface Registration. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007. 114

[15] Hamid Sarbazi-Azad in Hamid Sarbazi-Azad, Stupid Sort: A new sorting algorithm, Department of Computing Science Newsletter, University of Glasgow, 599:4, 2 October 2000.

[16] Gnome Sort , Levitin, Anany. Introduction to The Design & Analysis of Algorithms. Pages 69 – 84. Boston: Pearson Education Inc, 2003.

[17] T. Yamaguchi, S. Hashimoto. Image processing based on percolation model. IEICE Trans. Info. and Sys., E89- D(7): 2044–2052, 2006.

[18] A New Diagonal Hessian Algorithm for Blind Signal Separation, Elsabrouty, M.; Aboulnasr, T.; Bouchard, M. Sch. of Inf. Technol. & Eng., Ottawa Univ., Ont. , Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on May 2006

[19] Nafaa N, Latifa H, Djemel Z. Statistical tools for weld defect evaluation in radiographic

[20] Testing. European Conference on Non-destructive Testing 2006 Proceedings, Mo.2.4.4.

[21] [19]. Romeu R. da Silva, Marcio H.S. Siqueira, Marcos Paulo Vieira de Souza, João M.A. Rebello and Luiz P. Calôba. Estimated accuracy of classification of defects detected in welded joints by radiographic tests. NDT & E International, Volume 38, Issue 5, July 2005, Pages 335-343.

◈◈◈

The Challenges of Effectively Anonymizing Network Data

Ch. Himabindu, Y. Chitti Babu & P. Harini
Department of Computer Science Engineering, St Anns College of Engineering & Technology,
Chirala, 523115, India

1. Introduction

The availability of realistic network data plays a significant role in fostering collaboration and ensuring U.S. technical leadership in network security research. Unfortunately, a host of technical, legal, policy, and privacy issues limit the ability of operators to produce datasets for information security testing. In an effort to help overcome these limitations, several data collection efforts (e.g., CRAWDAD[14], PREDICT [34]) have been established in the past few years. The key principle used in all of these efforts to assure low-risk, high-value data is that of trace anonymization—the process of sanitizing data before release so that potentially sensitive information cannot be extracted.

Recently, however, the utility of these techniques in protecting host identities, user behaviors, network topologies, and security practices within enterprise networks has come under scrutiny. In short, several works have shown than unveiling sensitive data in anonymized network traces may not be as difficult as initially thought. The na¨ıve solution to this problem is to address the specifics of these attacks as they are discovered. However, doing so fails to address the underlying problem in its entirety. While isolated advances in network data anonymization are important, without a holistic approach to the problem they will simply shift the information-encoding burden to other properties of the traces, resulting in future privacy breaches. Given the significant reliance on anonymized network traces for security research, it is clear that a more exhaustive and principled analysis of the trace anonymization problem is in order.

Luckily, the problem of anonymizing publicly released data is not new. Over the past several decades, statisticians and computer scientists have developed approaches to anonymizing various forms of microdata, which are essentially databases of attributes collected about individuals. One prominent example is census data, which collects information about the salary, marital status, and many other potentially sensitive attributes from the population of an area or country. This census microdata, much like network data, is valuable to researchers for tracking trends, and as such the anonymized microdata must provide accurate information about potentially sensitive information. At the same time, it is essential that specifics from the data cannot be linked to individuals. In response, several anonymization methods, privacy definitions, and utility metrics have been developed to ensure that researchers can use the microdata for a wide spectrum of analyses while simultaneously providing principled, concrete guarantees on the privacy of those individuals within the data.

At first glance, it would seem as though the accumulated knowledge of microdata anonymization can be directly applied to network data anonymization since the two scenar ios share so much in common, including similar privacy and utility goals. Unfortunately, the inherently complex nature of network data makes direct application of these microdata methods difficult, at best. We can, however, learn from existing microdata anonymization literature and glean significant insight into how to approach the problem of network data anonymization in a principled fashion.

In this extended abstract, we compare and contrast the fields of microdata and network data anonymization to reveal the ways in which existing microdata literature may be applied to the network data anonymization problem. We further lay out several challenges that lie ahead in the development of robust network data anonymization methodologies that are grounded in the insights and lessons learned from microdata anonymization. Specifically, we examine the difficulties of clearly defining the privacy properties of network data due to its complex nature. In addition, we point out the necessity of utility measures in quantifying the extent to which anonymization may alter results obtained from analysis of the data. It is important to note that there are additional challenges that we do not address here, such as the legal and ethical issues with collecting network data. As a whole, we hope that this comparison between the fields of microdata and network data anonymization serves to fo-

cus the attention of the research community on a holistic approach to network data anonymization that enables the type of collaboration necessary to further progress in the areas of network and computer security research.

2. Microdata Anonymization

Roughly speaking, microdata can be thought of as a database with n rows and m columns, where each row in the microdata corresponds to a single entity that contributed its data. In the case of census data, for example, the rows might represent people who responded to the survey. The columns represent the attributes of those entities, such as their height or salary information. The goal of microdata anonymization is to alter the original data such that it is difficult (and quantifiably so) to infer potentially sensitive information about entities within the data while simultaneously ensuring that statistics computed on the data remain valid. As an example, average salary information for a given area should remain unchanged, but it should not be possible to infer a specific person's salary.

Specifically, the attributes of the microdata are divided into three categories: (i) identifiers, (ii) key attributes (i.e., quasi-identifiers), and (iii) sensitive attributes. Identifiers are attributes that trivially identify the row, such as name or social security number. Key attributes can be used to make inferences on the identity of the row from auxiliary sources of information. Though these key attributes do not directly identify the row, unique attribute values can be used to link rows in the anonymized microdata with other databases that do have identifying information. For instance, if a row in the microdata had a unique combination of height, weight, and age key attributes, then the adversary could use these attributes to look up the row's identity in a secondary database that includes the height, weight, age, and name. Finally, sensitive attributes are those that are not available from other data sources, and which the adversary would like to link to specific identities. To achieve the goal of anonymization, the data publisher removes identifiers, and applies one or more anonymization methods to alter the relationship between the key attributes and sensitive attributes to ensure that such inferences are unlikely. The resultant sanitized microdata can then be measured to quantify its level of privacy and utility.

2.1. Anonymization Methods

Several techniques are used by data publishers to anonymize microdata for publication. Truncation methods remove or reorganize records in the microdata to hide the relationship between the key attributes and sensitive attributes. These methods include removing rows, removing attributes, suppression of key attribute values in specific rows, or generalization (i.e., recoding) where several key attributes are combined into a single equivalence class (e.g., 25 _ age _ 35) [39]. Additionally, several methods based on perturbation of the sensitive attributes exist. Some examples of perturbation include swapping the values of sensitive attributes among different rows [15], sampling the data, or adding noise to the values [4, 17].

In addition to the truncation and perturbation-based methods, two methods have have been proposed which do not directly sanitize the microdata, but instead provide the underlying statistics of the data in alternate ways. The first of which, synthetic data generation [35, 28], attempts to model the original data and generate completely new microdata from that statistical model. Since this new data is generated from a model, the resultant microdata has no connection to real individuals and at the same time the specific statistical properties captured by the model are guaranteed to be preserved. The second method stores the data on a secure remote server, where the data user can access it only through a query interface [3, 40, 19]. Thus, the user only gets the answer to specific queries, and the query interface ensures that no queries are answered if they are harmful to privacy.

2.2. Measuring Privacy

Obviously, na¨ıvely applying anonymization methods to the data is not enough to guarantee privacy. In fact, inappropriate application of anonymization methods may provide several avenues of information leakage. For instance, a recent study by Narayanan and Shmatikov [30] showed that an anonymized dataset of movie recommendations released by NetFlix fails to meet the accepted privacy definitions for microdata, which results in re-identification of several users in the data. To prevent such information leakage, it is necessary to concretely measure the privacy of the resultant anonymized data. As the extensive literature in microdata privacy measures indicates, however, developing privacy definitions that encapsulate all areas of information leakage is not as straightforward as one might hope.

A common microdata privacy definition, known as k-anonymity, was proposed by Samarati and Sweeney [39]. The definition quantifies the difficulty of an adversary in determining which row in the microdata belongs to a given identity by requiring that every row must look like at least k _ 1 other rows with respect to their key attributes. In effect, this creates equivalence classes of key attributes where the adversary would have a 1=k chance of identifying the correct row using the key attributes. Chawla tal. [10] provide a similar notion of anonymity that applies to microdata containing numerical, rather than categorical, data types.

The notion of k-anonymity provides a necessary, but not Sufficient, condition for privacy since without it a row can be trivially identified by the uniqueness of its key attributes. Further restrictions are necessary, however, when we want to prevent the inference of sensitive attributes and not just which rows belong to a given identity. It may be possi-

ble, for example, to have an equivalence class that meets the k-anonymity definition, and yet has only one or a small number of distinct sensitive values. Thus, any individual that falls into such a class will have their sensitive attributes revealed. Machanavajjhala et al. [26]proposed `-diversity to strengthen the k-anonymity property by requiring that each class have at least Distinct sensitive values. Truta and Vinay [43] concurrently developed PSensitive Kanonymity to provide the same requirement.

The `-diversity property was further strengthened by Li et al. [25] since it may still be possible to leak information (in an information theoretic sense) about the sensitive attributes for an individual if the distribution of sensitive values in that individual's equivalence class is significantly different than those of the population. Essentially, the distribution within the equivalence class gives the adversary a more refined distribution of potential sensitive values for an individual than the adversary would have access to without the anonymized microdata. The t-closeness property [25] requires that the distribution of sensitive values in all equivalence classes be within a distancet of the population distribution across all rows in the microdata. This property ensures that the data publisher has greater control over the amount of information the adversary can gain about sensitive values of the individuals in the equivalence classes, thought small values of t clearly have a deleterious effect on the utility of the data.

While k-anonymity and t-closeness provide controls over the information disclosed by the key attributes and sensitive attributes, respectively, there are still other avenues of information leakage which the adversary can take advantage of. Zhang et al. [45] recently showed that it is possible to reverse the anonymization of a dataset if the adversary has knowledge of the anonymization method used (e.g., generalization). The key observation is that anonymization proceeds deterministically from anonymizations with the best utility (e.g., minimal equivalence class sizes) to those with worse utility, and will stop at the first anonymization that meets the privacy definition. Zhang et al.suggest the notion of P-safe,p-optimal anonymization, where anonymized microdata produced to meet privacy definition p (e.g., k-anonymity) is considered safe if it has more than one potential original microdata that could have produced it.

An alternative approach to these uncertainty, or indistinguishability, definitions is provided by the notion of differential privacy [16]. Differential privacy is primarily applied to interactive query systems where users interact with the data via a secure query interface. The notion of differential privacy states that the probability of a privacy breach occurring for a person is similar whether or not that person's information is contained in the data. The primary difference between differential privacy and the uncertainty-based definitions is that differential privacy is unable to quantify exactly what sensitive information could be leaked by the data, and instead focuses on the slightly more general guarantee that no additional harm will be done by adding a record.

2.3. Measuring Utility

The primary motivation for publishing anonymized microdata is to provide some utility, such as the ability to calculate statistics on the attributes, to researchers who make use of the data. Clearly, the data would be useless if the privacy definitions above are achieved at the expense of utility. As a result, several utility measures have been developed to provide researchers with metrics that allow them to gauge the confidence they should have in the results gained by analysis of the anonymized data. Most utility measures for microdata focus on specific utilities that are meant to be preserved. The obvious problem is that in doing so one can only anticipate a limited set of utilities and therefore can not offer guidance about other uses of the data.

Recently, some global utility measures have been proposed to try and quantify a wide range of utilities in a single metric [44, 20]. These global measures, however, can be difficult to interpret and often times do not strongly predict the available utilities. Specifically, these measures are loosely correlated with the extent to which utility is preserved, but they are unable to communicate to the researcher the exact way in which a particular utility is affected by the anonymization. For instance, Karr et al.'s use of the Kullback-Leibler divergence [20] between the anonymized and original data provides a broad notion of the similarity of the two distributions of attribute values, but that value has no direct connection to the changes to specific utilities.

3. Network Data Anonymization

Network data can be viewed in much the same way as microdata; containing n rows each representing a single packet (or summarized network flow) and columns representing the fields in the packet. Unlike microdata, which generally contains only categorical or numerical data, network data contains a variety of data types that make application of well-known anonymization methods difficult, if not impossible. Some fields in network data, like IP addresses, have a highly complex hierarchical ordering structure that often needs to be preserved after anonymization. Moreover, the relationship among different fields in network data is semantically rich, which means that the values taken by certain fields is dependent on their context with respect to other values within the data – both within the same row and within other rows – and these dependencies must be maintained in order for the data to be semantically meaningful.

The goals of network data anonymization are also superficially similar in nature to those of microdata insofar as they are focused on preventing the disclosure of sensitive information about certain entities present within the

data. However, these goals are far more nebulous in the network data case since this sensitive information cannot be defined as a single field, nor can it be quantified for just a single row. Network data publishers are concerned with the privacy of workstations on the network and their users, which can be associated with multiple rows (e.g., packets) within the data. The sensitive information about these entities is often encoded in complex relationships among multiple fields across several different rows, such as a user's web browsing patterns or computer virus activity. Unfortunately, these goals remain ill-defined even in the most recent work in this area, which necessarily limits the efficacy of the anonymization procedures.

3.1. Anonymization Methods

Currently, the anonymization of network data is performed by applying one of a limited number of techniques, many of which are shared with microdata, to fields in the data chosen by the data publisher and defined in an anonymization policy language [33, 42]. The most widely used of these techniques are truncation, randomization, quantization, and pseudonymization. Truncation and randomization effectively destroy the semantics of the field they are applied to, but are helpful when dealing with fields that are likely to contain highly sensitive data. One example is the payload of packets, which might contain usernames and passwords and are removed from the data as standard practice. Quantization techniques, such as limiting the precision of time stamps, are applied to reduce the information gained about the identity of the workstations from timing attacks [21]. Perhaps the most widely used technique, pseudonymization, replaces IP addresses found in the data with linkable, prefix-preserving pseudonyms [32, 18]. These pseudonyms preserve the hierarchical relationships found in the prefixes of the original addresses. The underlying goal is to enable the analysis of packets generated from hosts, or whole prefixes, without providing the actual IPs.

In an effort to maintain as much of the original data as possible, data publishers apply these methods to as few fields as possible; normally, just the IP addresses, time stamps, and payloads. In fact, fields within the network data are typically anonymized only when they are shown to leak information via published attacks. As a result, the unaltered fields of the data provide significant information that can be used as key attributes to link objects in the data to their real identities. This reactionary anonymization policy has lead to the discovery of several attacks which use the unaltered features of the data to re-identify workstations and their behaviors [37, 5, 6, 12], and identify web pages that the users visit [22, 11].

3.2. Measuring Privacy

Given the reactionary nature of network data anonymization, it comes as no surprise that network data does not have well-defined privacy measures, due in part to the difficulty in clearly defining the privacy properties desired by data publishers. To date, there have been a few attempts to quantify the uncertainty that the adversary has in identifying which pseudonyms or values in the data belong to which real world workstations. For instance, Ribeiro et al. [37] derive fingerprints, such as the port numbers used, for each IP address in both the anonymized and original data, and compare the two sets of fingerprints to determine the equivalence classes for each IP address. Those workstations with highly unique fingerprints are considered to be privacy risks for the data publisher. Coull et al. [13] also examines the similarity between the anonymized and original data, but examines a broader range of distributions of values found in the data. In doing so, they quantify the privacy of workstations in terms of the number of other workstations with similar value distributions, and also discover those fields in the data that negatively affect privacy. Kounine and Bezzi [23] perform a similar analysis with respect to the privacy of individual values after they have been anonymized rather than workstation privacy as a whole. The problem, of course, is that each of these techniques focus exclusively on workstation or individual field privacy, and yet network data can contain several different types of entities whose privacy is equally important.

3.3. Measuring Utility

The idea of quantifying the utility of network data is only just beginning to gain traction in the network data anonymization community, though the complex nature of the data makes such measures as important, if not more so, as those proposed in microdata anonymization. One such utility measure was recently proposed by Lakkaraju and Slagell [24], and compares the performance of a well-known intrusion detection system on the anonymized and unanonymized data. Another measure was proposed by Burkhart et al. [8] and applies anomaly detection methodologies to the anonymized data to quantify the way in which it affects its performance. Both methods closely resemble those of Brickell and Shmatikov [7] that apply machine learning tasks to microdata to determine the degradation in accuracy. In addition, the global utility measure of Woo et al. [44] can also be adapted to network data due to its use of standard statistical classification techniques. As with microdata, the use of highly specific measures, such as evaluations under specific anomaly detection methodologies or intrusion detection systems, leads to results that may not be applicable in a more general context. Similarly, global measures still remain difficult to interpret due to their disconnection from concrete utilities, and may in fact be even

more difficult to apply effectively to network data because of its inherently complex and interdependent nature.

4. The Challenges Ahead

Clearly, the problem of anonymizing microdata has received significant attention over the past three decades, and that attention has served to develop several methodologies for providing private and useful microdata to researchers. It is equally clear that network data anonymization is only just beginning to mature as an area of active research, and it can benefit from the substantial body of work generated by microdata anonymization research due to the similarities between the two areas. That said, microdata and network data have a number of non-trivial differences that make direct application of well-known microdata anonymization concepts meaningless. In this section, we outline three broad challenges that lie ahead in the development of effective methods for anonymizing network data.

4.1. What are we protecting?

Before we can begin developing effective anonymization methods for network data, we must first have a clear understanding of exactly what it is we hope to protect. For microdata, this question is easily answered because there is a natural one-to-one correspondence between the rows in the data and the entities being protected. With network data, however, this connection is not as clear. Publishers of network data are interested in protecting the privacy of a number of entities: the network users, the network's security procedures, and the hosts that operate on the network. What makes it difficult to clearly define these entities is the fact that network data is inherently multifaceted. A single record in the network data may actually affect the privacy of many entities of varying types. Moreover, the privacy of those entities is not contingent on only a single row in the data, but on many rows that define their behavior over time. These issues naturally raise questions about how we define each of the entities for which the data publisher is interested in providing privacy.

With that said, for some types of entities the answer to this question is relatively straightforward. When considering the privacy of hosts on the network, for example, these host entities can be defined by assuming that the IP addresses in the network data consistently and uniquely identify a host. Even so, the relatively simple entity definition of hosts is not without its caveats, such as the possibility that multiple hosts may use the same IP. More complex entities, like users or web pages, are more difficult to define without significant auxiliary information (e.g., audit logs). Using those auxiliary data sources to mark the entities associated with each record in the data is one potential avenue for defining the entities of interest in the network data.

4.2. What is sensitive?

Network data has a wide variety of information encoded within it. One need only consider some of its uses in network research to appreciate its scope: e.g., measurements of network traffic characteristics, testing new networking methodologies and tools, and studying emerging phenomena. As we move forward, we must decide which of these pieces of information encoded within the network data should be considered to be sensitive. Again, the relatively simple structure of microdata allows for an elegant definition of sensitive information – any attribute in the data that is likely to be unavailable from an external information source should be labeled as sensitive. The sensitivity of attributes are often easily intuited from knowledge of the underlying data. Unfortunately, such intuitive definitions are simply not applicable to network data.

The very same information-rich properties that make network data so useful to the research community also lead to two significant challenges in defining which pieces of information might be considered sensitive. First, potentially sensitive information encoded within the network data is not restricted to a single column in the data. In fact, the relationships between the columns and across several records often indicate the most sensitive of information. For instance, the distribution of ports used by a host in combination with other fields may indicate that the host is infected by a virus, whereas the distribution of ports alone would not. Similar arguments could be made for whether a user visited an illicit web site, or if the network is using a particular security system. Second, many of the fields present within network data contain a combination of both publicly known and private values. As an example, the distribution of ports used by a host may indicate the services it offers, both publicly and privately within the local network. These scenarios are particularly troublesome since the known values within the column of port numbers can act as key attributes, while the unknown values act as sensitive attributes that the adversary may seek to infer.

Many of the attacks that have been discovered for anonymized network data take advantage of these issues in subverting the privacy of the data. Host profiling attacks [37, 5, 6, 12], for instance, use some of the ports and IP pseudonyms in the data as key attributes to link the hosts to their real identities, and then use the remaining ports to infer the hosts hidden services. Rather than attempt to adapt the static notions of key and sensitive attributes to multifaceted network data, current approaches to measuring privacy of network data (e.g., [37, 13, 23]) instead focus on the uniqueness of a piece of data as an indicator for sensitivity. The underlying assumption is that a sufficiently unique behavior encoded within the data is likely to be unavailable from other data sources.

4.3. Defining Utility for Network Data

An area of considerable interest for both microdata and network data anonymization is the development of metrics that measure the utility of the data after it has been anonymized.These metrics are especially important in the case of network data, where the inherent difficulties of defining sensitivity and entities within the data may lead to essentially useless data. For instance, if we follow the definition that sensitive information in network data is any piece of information that is sufficiently unique, then it is easy to imagine a scenario in which the network data contains only homogenous behavior. This type of homogenous data would be useless to researchers who are interested in investigating anomalous incidents or who want to get an accurate estimation of traffic characteristics. In these types of scenarios, it is imperative that researchers have access to utility metrics with respect to certain properties of the data so that they, and those that review their work, can adequately gauge its appropriateness to the task at hand.

Specific utility measures may provide an adequate short term solution to the problem. In general, a utility measure can be derived by comparing the results of a particular utility on the anonymized data to those of the unanonymized data. The problem, of course, lies in predicting the utilities that will be used. One simple way to address this concern is for the data publisher to publish a set of metrics for standard utilities on the data, and allow researchers to request additional utility measures as necessary. However, this type of arrangement is a significant burden on data publishers and researchers, since data publishers would need to run various analyses on the data and researchers would be unable to perform exploratory analyses in a timely fashion. A slightly different approach might be to adapt the concept of a remote verification server, such as the one proposed by Reiter et al. [36], to allow researchers to automatically compare their results from the anonymized data with those from the original data with respect to a specific utility.

5. CONCLUSION

The uncertainties that currently exist about the efficacy of network data anonymization, from both technical and policy perspectives, leave the research community in a vulnerable position. Even as the field marches forward, it does so with little understanding of the implications of publishing anonymized network data on the privacy of the networks being monitored and the utility to researchers. Without that understanding, data publishers are left to wonder what fields must be anonymized to avoid legal fallout, while researchers question the confidence of results gained from the data. However, the extensive work done on microdata anonymity provides the network research community with several useful insights about how to effectively apply anonymization to published data. At the same time, this prior wisdom cannot be applied directly without first overcoming several challenges, including the development of appropriate privacy and utility definitions for the more complex case of network data. Addressing these challenges is essential, in our view, to ensure the continued, yet responsible, availability of network trace data to support security research.

REFERENCES

[1] M. Allman and V. Paxson. Issues and Etiquette Concerning Use of Shared Measurement Data. In Proceedings of the ACM SIGCOMM Internet Measurement Conference , page To appear, 2007.

[2] R. J. Bayardo and R. Agrawal. Data Privacy through Optimal k-Anonymization. In Proceedings of the 21st International Conference on Data Engineering , pages 217˜, 2005.

[3] A. Blum, C. Dwork, F. McSherry, and K. Nissim. Practical Privacy: The SuLQ Framework. In Proceedings of the 24th ACM Symposium on Principles of Database Systems , pages 128`, 2005.

[4] R. Brand. Microdata Protection through Noise Addition. In Inference Control in Statistical Databases, From Theory to Practice , pages 97N, 2002.

[5] T. Brekne and A. °Arnes. Circumventing IP-Address Pseudonymization. In Proceedings of the 3R d IASTED International Conference on Communications and Computer Networks, pages 43¨, October 2005.

[6] T. Brekne, A. Arnes, and A. Øslebø. ° Anonymization of IP Traffic Monitoring Data – Attacks on Two Prefix-preserving Anonymization Schemes and Some Proposed Remedies. In Proceedings of the Workshop on Privacy Enhancing Technologies, pages 179Ž, May 2005.

[7] J. Brickell and V. Shmatikov. The Cost of Privacy: Destruction of Data-Mining Utility in Anonymized Data Publishing. In Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining , pages 70À, 2008.

[8] M. Burkhart, D. Brauckhoff, and M. May. On the Utility of Anonymized Flow Traces for Anomaly Detection. In Proceedings of the 19th ITC Specialist Seminar on Network Usage and Traffic, October 2008.

[9] A. J. Burstein. Conducting Cybersecurity Research Legally and Ethically. In USENIX Workshop on Large-scale Exploits and Emergent Threats, April 2008.

[10] S. Chawla, C. Dwork, F. McSherry, A. Smith, and H. Wee. Toward privacy in Public Databases. In

Proceedings of the 2nd Annual Theory of Cryptography Conference , pages 363–385, 2005.

[11] S. Coull, M. Collins, C. Wright, F. Monrose, and M. K. Reiter. On Web Browsing Privacy in Anonymized NetFlows. In Proceedings of the 16 thUSENIX Security Symposium, pages 339ê, August 2007.

[12] S. Coull, C. Wright, F. Monrose, M. Collins, and M. K. Reiter. Playing Devil's Advocate: Inferring Sensitive Information from Anonymized Network Traces. In Proceedings of Prothe 14th Annual Network and Distributed System Security Symposium, pages 35§, February 2007.

[13] S. E. Coull, C. V. Wright, A. D. Keromytis, F. Monrose, and M. K. Reiter. Taming the Devil: Techniques for Evaluating Anonymized Network Data. In Proceedings of the 15th Network and Distributed Systems Security Symposium,pages 125], 2008.

[14] CRAWDAD: A Community Resource for Archiving Wireless Data at Dartmouth. http://crawdad.cs. dartmouth.edu .

[15] T. Dalenius and S. P. Reiss. Data-swapping: A Technique for Disclosure Limitation. Journal of Statistical Planning and Inference, 6:73Å, 1982.

[16] C. Dwork. Differential Privacy. In Proceedings of the 33rd International Colloquium on Automata, Languages, and Programming (ICALP) , pages 1Š, 2006.

[17] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. Our Data, Ourselves: Privacy via Distributed Noise Generation. In Proceedings of Advances in Cryptology– EUROCRYPT, pages 486C, 2006.

[18] J. Fan, J. Xu, M. Ammar, and S. Moon. Prefix-preserving IP Address Anonymization: Measurement-based Security Evaluation and a New Cryptography-based Scheme. Computer Networks, 46B):263º, October 2004.

[19] S. Gomatam, A. F. Karr, J. P. Reiter, and A. P. Sanil. Data Dissemination and Disclosure Limitation in a World Without Microdata: A Risk-Utility Framework for Remote Access Analysis Servers. Statistical Science, 20B):163, 2005.

[20] A. F. Karr, C. N. Kohnen, A. Oganian, J. P. Reiter, and A. P. Sanil. A Framework for Evaluating the Utility of Data Altered to Protect Confidentiality. The American Statistician , 60C):224š, 2006.

[21] T. Kohno, A. Broido, and K. Claffy. Remote Physical Device Fingerprinting. In Proceedings of the IEEE Symposium on Security and Privacy , pages 93H, May 2005.

[22] D. Koukis, S. Antonatos, and K. Anagnostakis. On the Privacy Risks of Publishing Anonymized IP Network Traces. In Proceedings of Communications and Multimedia Security ,pages 22š, October 2006.

[23] A. Kounine and M. Bezzi. Assessing Disclosure Risk in Anonymized Datasets. In Proceedings of FloCo, n 2008.

[24] K. Lakkaraju and A. Slagell. Evaluating the Utility of Anonymized Network Traces for Intrusion Detection. Inceedings of the 4th Annual Conference on Security and Privacy in Communication Networks , .September 2008.

[25] N. Li, T. Li, and S. Venkatasubramanian. t -Closeness: Privacy Beyond k-Anonymity and l -Di versity. In Proceedings of the 23rd IEEE International Conference on Data Engineering, pages 106M, April 2007.

[26] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam. l -Di versity: Privacy Beyond k -Anonymity. In Proceedings of the 22 nd IEEE International Conference on Data Engineering , pages 24ı, April 2006.

[27] J. M. Mateo-Sanz, J. Domingo-Ferrer, and F. Seb´e. Probabilistic Information Loss Measures in Confidentiality Protection of Continuous Microdata. Data Mining and Knowledge Discovery , 11B):181 ‹, 2005.

[28] J. P. Mateo-Sanz, A. Martinez-Balleste, and J. Domingo- Ferrer. Fast Generation of Accurate Synthetic Microdata. In Proceedings of the International Workshop on Privacy in Statistical Databases , pages 298Æ, 2004.

[29] J. Mirkovic. Privacy-Safe Network Trace Sharing via Secure Queries. In Proceedings of the 1st ACM Workshop on Network Data Anonymization , October 2008.

[30] A. Narayanan and V. Shmatikov. Robust De-anonymization of Large Sparse Datasets. In Proceedings of the 2008 IEEE Symposium on Security and Privacy , pages 111U, 2008.

[31] P. Ohm, D. Sicker, and D. Grunwald. Legal Issues Surrounding Monitoring During Network Research (Invited Paper). In ACM SIGCOMM/USENIX Internet Measurement Conference, San Deigo, October 2007.

[32] R. Pang, M. Allman, V. Paxson, and J. Lee. The Devil and Packet Trace Anonymization. ACM Computer Communication Review, 36A):29 , January 2006.

[33] R. Pang and V. Paxson. A High-Level Environment for Packet Trace Anonymization and Transformation. In Proceedings of the ACM Special Interest Group in Communications (SIGCOM) Conference , pages 339é, August 2003.

[34] PREDICT: Protected Repository for the Defense of Infrastructure Against Cyber Threats. http://www.predict. org .

[35] T. E. Raghunathan, J. P. Reiter, and D. B. Rubin. Multiple Imputation for Statistical Disclosure Limitation. Journal of Official Statistics , 19:1Ž, 2003.

[36] J. P. Reiter, A. Oganian, and A. F. Karr. Verification Servers: Enabling Analysts to Assess the Quality of Inferences from Public Use Data. Computational Statistics and Data Analysis, forthcoming.

[37] B. Ribeiro, W. Chen, G. Miklau, and D. Towsley. Analyzing Privacy in Enterprise Packet Trace Anonymization. In Proceedings of the 15 th Network and Distributed Systems Security Symposium, to appear , 2008.

[38] D. B. Rubin. Satisfying Confidentiality Constraints Through the Use of Synthetic Multiply-Imputed Microdata. Journal of Official Statistics , 9:4618, 1993.

[39] P. Samarati and L. Sweeney. Protecting Privacy When Disclosing Information: k-Anonymity and its Enforcement Through Generalization and Suppression. Technical Report SRI-CSL-98-04, SRI Computer Science Laboratory, 1998.

[40] B. Schouten and M. Cigrang. Remote Access Systems for Statistical Analysis of Microdata. Statistics and Computing , 13D):381°, 2003.

[41] C. Shannon, D. Moore, and K. Keys. The Internet Measurement Data Catalog. ACM SIGCOMM Computer Communications Review, 35E):97@, Oct. 2005. See http: //imdc.datcat.org/browse .

[42] A. Slagell, K. Lakkaraju, and K. Luo. FLAIM: A Multilevel Anonymization Framework for Computer and Network Logs. In Proceedings of the 20 th USENIX Large Installation System Administration Conference, pages 63¿, 2006.

[43] T. M. Truta and B. Vinay. Privacy Protection: p-Sensitive k-Anonymity Property. In Proceedings of the 2nd International Workshop on Privacy Data Management , page 94, 2006.

[44] M. Woo, J. P. Reiter, A. Oganian, and A. F. Karr. Global Measures of Data Utility in Microdata Masked for Disclosure Limitation. Journal of Privacy and Confidentiality , forthcoming.

[45] L. Zhang, S. Jajodia, and A. Brodsky. Information Disclosure Under Realistic Assumptions: Privacy versus Optimality. In Proceedings of the 14th ACM Conference on Computer and Communications Security , pages 573ƒ, 2007.

◈◈◈