CVR JOURNAL OF SCIENCE & TECHNOLOGY

CVR COLLEGE OF ENGINEERING

(UGC Autonomous - Affiliated to JNTU Hyderabad)
Mangalpalli (V), Ibrahimpatnam (M),
R.R. District, Telangana. – 501510
http://cvr.ac.in

EDITORIAL

We are happy to bring out Volume 8 of the Biannual Journal of our college, CVR Journal of Science & Technology, with a wide spectrum of papers from different branches of Engineering, Science and Management. As was the case for the previous volumes, we received a number of papers. Out of the papers received, a limited number of 17 papers were selected for inclusion in the current volume, based on the recommendations made by the members of the Editorial Committee.

The breakup of the papers among the various disciplines is as follows:

CSE - 1; ECE - 5; EEE - 2; EIE - 3; IT - 2; H&S - 3; Mgmt -1

The authors include staff members of our college and some from external institutions as coauthors.

The Computer Science paper covers Spatiotemporal Data Indexing Structures.

Topics selected for publication in ECE include research work into advanced subjects like cognitive radio, spectrum sensing, OFDM system, higher order modulation schemes and wavelets etc.

The papers that are selected from EEE cover current trends and include topics such as cascaded multilevel inverters and microgrid.

We hope that the interest of staff members of all the departments of our college will grow further in contributing their research work in the journal. We have been including some papers from external organizations and it will be our attempt to attract more numbers from outside so as to broaden the scope of our journal.

I am grateful to all the members of the Editorial Board for their help in short listing papers for inclusion in the current volume of the journal. I wish to thank Dr. K. Venkateswara Rao, Professor of CSE and Sri S.Prashanth, DTP Operator in the Department of H&S for help in the preparation of the papers in camera ready form for final printing.

K.V.Chalapati Rao
Editor

CONTENTS

Review of Spatiotemporal Data Indexing Structures

K.Venkateswara Rao[1], A.Govardhan[2] and K.V.Chalapati Rao[3]
[1,3]Department of CSE, CVR College of Engineering, Ibrahimpatnam, Hyderabad, India
[1]kvenkat.cse@gmail.com
[3]chalapatiraokv@gmail.com
[2]School of Information Technology, JNTUH Hyderabad, India

govardhan_cse@yahoo.co.in

Abstract: **Spatiotemporal data are those objects that change their position and/or extent over time. There are many applications like geographic information systems, intelligent transportation systems, environmental information systems and mobile communications data management systems which deal with spatiotemporal data. Efficient processing of the spatiotemporal data needs sophisticated indexing schemes. Many indexing structures are available in the literature for accessing such data sets. This paper provides a comprehensive survey, classification and brief description of spatiotemporal data indexing structures for efficient accessing of the data by various types of spatiotemporal queries.**

Index terms: ***Database index, spatiotemporal index, aggregate data index***

I. INTRODUCTION

A spatiotemporal object can be represented as (NScT, ScT, T) [29] where NScT is the time-stamped non-spatial conventional component, ScT is the time-stamped spatial component and T is the temporal component indicating its valid time. The spatial and non-spatial conventional components can take many values along time line. ScT is a set of records where each record contains a spatial component value at i^{th} time point, Ti. Similarly NScT is a set of records where each record contains non-spatial component value at time point, Ti. All Ti's are contained in T and they do not intersect.

The mathematical model defining spatiotemporal data structure is given below.

SDS = (THC, SC, TC) where
SDS = Spatiotemporal Data Structure
THC = Thematic Component
SC = Spatial Component
TC = Temporal Component
THC = { (A,T) /A = Thematic Attribute Л T = Temporal Attribute }
SC = { (S,T) /S = Spatial Attribute Л T = Temporal Attribute }
TC = { T / T = Temporal Attribute }

Thematic Attributes are non spatial properties of spatial objects whose values may change with time. Spatial Attributes are geometry and location of the spatial objects whose values also may change with time. Temporal Attributes are temporal point and temporal interval.

The aim of indexing spatiotemporal data is to facilitate its retrieval for efficient processing of spatiotemporal queries such as time slice queries e.g., " Find all objects that cross a certain area at time t", window queries or range queries e.g., " Find all objects that cross a certain area in the time interval [t1,t2]", n-nearest neighbor queries e.g., " Find n hospitals that are closest to a given moving point", trajectory queries e.g., " Find trajectory of a given point for the past one hour". More research is done in developing spatiotemporal data indexing and access methods to support spatiotemporal queries. The spatiotemporal access methods [1,2] with underlying spatial and temporal structures are grouped into different categories as follows.

- Indexing the past data
- Indexing the current data
- Indexing the future data
- Indexing data at all points of time
- Indexing aggregate historical spatiotemporal data
- Miscellaneous indexing structures

This paper is organized as follows. Section 1 presents definition of spatiotemporal object and classification of spatiotemporal data indexing structures. Review of the literature relevant to Indexing the past data is given in Section 2. Indexing the current spatiotemporal data related work are reviewed in Section 3. Section 4 deals with indexing structures for the future spatiotemporal data. Section 5 describes an Indexing Past, Present and Future Positions of the spatiotemporal data. Indexing structures for aggregate historical spatiotemporal data are reviewed in Section 6. Other miscellaneous indexing structures for spatiotemporal data are provides in section 7. Finally section 8 provides conclusion

II. INDEXING THE PAST

The spatiotemporal indexing methods for historical data are grouped into three categories.

- First one is dealing with temporal dimension.
- Second one is dealing with overlapping and Multi-version structures
- Third category is dealing with Trajectory oriented access methods.

2.1 Indexing Structures dealing with Temporal Dimension

In this category, the main focus is to deal with the spatial domain. The temporal dimension is handled by augmenting the temporal aspect into existing spatial access methods. RT-tree, 3DR-tree, STR-tree, MTSB-tree, FNR-tree and MON-tree fall into this category. RT-tree [1] combines R-tree [3] as spatial access method and Time-Split B-tree (TSB-tree) [1] as a method for temporal access. It indexes entire spatiotemporal data in a single R-tree [3]. It is very difficult to manage RT-tree when more number of objects change and interval queries span entire tree. 3DR-tree [1,4,5,6] handles interval queries efficiently by treating time as third dimension but its performance is poor on timestamp queries. STR-tree [1,7] is a variant of R-tree. R-tree is extended with a different algorithm for inserting and splitting. The Multiple TSB (MTSB) tree [2,8] uses a TSB-tree for supporting spatial range and historical queries on a database of moving objects. Objects moving in a set of connected and fixed line segments in a two-dimensional space are indexed by Fixed Network R-tree (FNR-tree) [2,9]. FNR-tree is extended by MON-tree [2,10]. The constrained network is modeled in MON-tree as a set of junctions and routes that are non-intersecting poly-lines.

2.2 Overlapping and Multi-version Structures

In this category, the temporal aspects are separated from spatial aspects so that entire spatial data that is valid at one time instant is kept together and managed in one structure. Excessive storage is required by this approach. Consecutive instances are combined into a single structure using overlapping in access methods to avoid storage of identical sub-structures. MR-tree, HR-tree, HR+-tree, MV3R-tree, PA-tree and GS-tree belong to this category. Both MR-tree [1] and Historical R-tree (HR-tree) [4,5,11] create a separate R-tree to every timestamp. This allows sharing the branches of consecutive trees to avoid storage overhead. Both these structures are efficient for timestamp queries but their performance on interval queries is poor. HR+-tree [1,11] is developed to avoid the replication problem in the HR-tree. Multi-version 3DR-tree (MV3R-tree) [1,4] mainly depends on multi-version B-tree (MVB-tree) [4]. It combines a multi-version R-tree (MVR-tree) [4] and a 3DR-tree so that timestamp queries are processed using MVR-tree [4,12] and interval queries are processed using 3DR-tree [4,5,6]. Partially persistent R-tree (PPR-tree) [1,11,13] structure is designed for bi-temporal databases. It maintains R-tree evolution so that any historical query at time point t is able to use R-tree structure corresponding to time t. The trajectories of moving objects are indexed using Parametric tree (PA-tree) [2,14], that indexes by splitting time duration of the moving object into n disjoint time intervals and also dividing trajectory of the object into n line segments corresponding to the respective time intervals. Graph Strip tree (GS-tree) [2,15] indexes historical position data of moving objects using a constrained graph that is defined by vertices and edges as in MON-tree.

2.3 Trajectory-Oriented Access Methods

This category of access techniques focuses on trajectory-oriented queries on historical spatiotemporal data. TB-tree, SETI, SEB-tree, CSE-tree, Polar-tree and RTR-tree fall into this category. Trajectory-bundle tree (TB-tree) [1,16] structure is similar to R-tree and is an extension of STR-tree [7]. A leaf node of a TB-tree stores line segments of same trajectory. The drawback is that line segments belonging to the trajectories of spatially co-located moving objects are maintained in different nodes of TB-tree. Scalable and Efficient Trajectory Index (SETI) [1] creates static, non-overlapping partitions for spatial dimension. The trajectory segments within each partition are indexed using R-tree. Start/End timestamp B-tree (SEB-tree) [1] divides the space into zones. It indexes each zone. The zones may overlap. A hash function that is based on both timestamps of a moving object is used to hash the object into its zone. Compressive Start-End Tree (CSE-tree) [2,17] divides the space into disjoint cells and maintains a time for each spatial cell. It is used in GPS data sharing applications. Polar-tree [2] is used to index orientations of moving objects to a given focal point to detect the objects which get closer or move away from the focal point. RTR-tree[2,18] is based on R-tree. It is used to index trajectories of objects which are moving in symbolic indoor space. The RFID readers are used to gather position data of the moving objects. HBSTR-Tree [19] is a hybrid index structure comprising spatio-temporal R-tree2, B-tree and hash table. It is used for trajectory databases.

III. INDEXING THE CURRENT POSITIONS

The idea of the "current" positions [1] is challenging in database systems. Spatiotemporal index methods that support queries involving current positions data are 2+3 R-tree, 2-3 TR-tree, LUR-tree, LUGrid, RUM-tree and IMORSS. The 2+3 R-tree [1] maintains two R-trees. First one is used to index two dimensional current positions of the moving object. Second one is used to index the trajectories in spatial and temporal dimensions to maintain the history of the moving objects. The trajectories are built for updated current objects in three dimensions. Then the objects are inserted into the second R-tree and deleted from the first R-tree. The 2-3 TR-tree [1] is similar to the 2+3 R-tree except that its three dimensional R-tree does not maintain the trajectories. It maintains only multi-dimensional points. It uses the underlying structure of TB-tree to answer trajectory-oriented queries. The Lazy Update R-tree (LUR-tree) [1] is only concerned with the current locations of the objects. It stores no historical data. The Lazy-Update Grid-based (LUGrid) [2,20] index uses grid file for lazy insertion and deletion. The positions of moving objects are updated frequently through the grid file. The current object identifiers are tracked in LUGrid using in-memory data structures. The R-tree with Update Memo (RUM-tree) [2,21] manages frequent updates to locations of moving objects by inserting the object in the new position without touching its old location. The obsolete entries in the RUM-tree are removed by lazy garbage cleaner. Indexing Moving Objects on Road Sectors (IMORS) [2,22] is used to index current locations

of moving objects on a fixed route network. R*-tree [3] is used in IMORS to index road sectors.

IV. INDEXING THE FUTURE POSITIONS

It is required to have extra data like destination and speed of a moving object to predict its future locations in a variety of ways such as the original space-time continuum, transformation methods and parametric spatial access methods.

4.1 The Original Space-time Continuum

The moving object in one dimensional space can be modeled using linear equation $p_t = vt + c$ where v is the constant velocity, c is a constant for the starting position of a moving object and p_t is the predicted location at time t. The predicted locations of moving objects can be computed and represented in the original spatiotemporal space. PMR-quadtree [23] and MOVIES [24] follow this approach. PMR-quadtree [23] for moving objects indexing destroys whole index structure and rebuilds it again when an update to moving objects occurs. In this way, it only maintains current PMR-quadtree. Moving objects indexing using frequent snapshots (MOVIES) [24] uses linearized kd-trees [3] to index locations of moving objects for supporting predictive queries on the moving objects.

4.2 Transformation Methods

These methods transform original time-space domain into another space representation to ease the representation and querying of the data in future. Duality transformation [1,25], SV-Model, PSI, STRIPES, Bx-tree, By-tree, ST2B-tree, Bdual-tree and BdH-tree are transformation methods based indexing techniques [2,1]. Duality transformation [1,25] represents the equation $p_t = mt + n$ as a point (m, n) in a dual two dimensional space in which horizontal dimension is the velocity m and vertical dimension is the reference location n. KD-tree [3] based spatial index is used instead of an R-tree in the dual space due to highly skewed nature of the distribution. SV-model [1] represents a moving object using four parameters - starting position, starting timestamp, initial velocity and destination. The horizontal dimension in the dual space representation is starting time and the vertical dimension is the destination. It assumes starting position and Velocity as constants. The dual space is indexed by SS-tree. PSI [1,26] approach uses an R-tree to index three dimensional space for modeling trajectories. The three dimensions correspond to reference location, velocity and time. STRIPES [2] models positions of a moving object using a linear function in n-dimensional space and predicted locations of the moving object in n-dimensional space are transformed into points in a two dimensional dual space. It maintains two different indexes using PR-quadtree. These two indexes correspond to first and second half of time intervals of the moving object. Bx-tree [2] is an extension of the B+-tree for indexing moving objects. By-tree [2] is an extension of the Bx-tree. It uses different update frequencies for each moving object. ST2B-tree [2] indexes the future positions of moving objects in a manner similar to Bx-tree. It improves Bx-tree by handling data skewness in both spatial and temporal dimensions. Bdual-tree [2] represents both d-dimensional positions and velocities in a dual two dimensional space. It is built on Bx-tree. BdH-tree [2] builds on Blink-tree instead on Bx-tree. The internal nodes of Blink-tree at the same level are linked.

The transformation methods have some drawbacks. The entire data in the primal space may not be captured in the dual space. It is not guaranteed that objects which are near in original space continue to be near in dual space.

4.3 Parametric *Spatial Access Methods*

Parametric bounding rectangles are used in these methods to index the original time-space. These rectangles are represented as functions of time in such a way that the enclosed moving object in the rectangle remains to be in the same rectangle. TPR-tree, PR-tree, VCI R-tree, STAR-tree, R^{EXP}-tree, TPR*-tree, STP-tree and ANR-tree fall into this category. Time Parameterized R-tree (TPR-tree) [1,27] constructs conservative bounding rectangles in R-tree. They enclose a set of moving points. PR-tree [1,28] is similar to TPR-tree except that it represents the moving object as spatial extents. Velocity Constrained Indexing R-tree (VCI R-tree) [1] imposes a restriction on maximum speed of moving objects. It is capable of addressing the problems in continuous queries execution. Spatio-Temporal Self Adjusting R-tree (STAR-tree) [1] is same as TPR-tree except that it is able to self adjust without user input when the query performance degrades. R^{EXP}-tree [1] has extended TPR-tree to use expiration times to manage moving objects. The expired entries are removed from the index using a lazy technique. TPR*-tree [1] is same as TPR-tree except that it uses insert and deletion algorithms of R*-tree. TPR-tree and TPR*-trees are generalized in Spatio-Temporal Prediction Tree (STP-tree) [2]. Future locations of moving objects are indexed using arbitrary polynomial type moving functions in STP-tree. Adaptive Network R-tree (ANR-tree) [2,29] is used to index future trajectories of objects moving in a constrained network.

V. INDEXING PAST, PRESENT AND FUTURE POSITIONS

R^{PPF}-Tree, $PCFI^{+}$ - index, BBx-index, UTR-tree, $STCB^{+}$-tree and PPFI index structures fall into this group. R^{PPF}-tree [2,30] indexes locations of moving objects at all timestamps. The past positions of a moving object are captured in R^{PPF}–tree by the application of partial persistence to the TPR-tree. The future locations are calculated using interpolation of the past locations of a moving object using a linear function. The time-parameterized bounding rectangle (TPBR) in R^{PPF}–tree is used for current time. The past-current-future-index (PCFI) [2,30,31] is built using TPR*-tree and SETI. Moving object space is split into non-overlapping cells in PCFI. Present locations and velocities of a moving object are indexed using TPR*-tree and is maintained for each cell. PCFI indexes historical data using a sparse on-disk R*-tree. BBx-index [30,32] uses Bx-tree to support the present and future. Uncertain Trajectory R-tree (UTR-tree) [2] is an extension of the MON-tree with uncertainty that belongs to a constrained network. Spatio-Temporal Compressed B+ - tree (STCB+ - tree) [2] indexes

trajectories of moving spatiotemporal objects using multiple compressed B+-trees. PPFI [2] indexes road network using 2D R*-tree. It uses hash structure to predict near-future locations and to describe the recent states of moving objects.

VI. Indexing Aggregate Historical Spatiotemporal Data

Many real-life applications need to access summarized spatiotemporal data. Pre-aggregation of raw data is required in these applications due to involvement of huge amount of spatiotemporal data organized in multidimensional data structure. Both spatial and temporal dimensions can be associated with dimension hierarchies. Queries require aggregation of the data along the dimension hierarchies to satisfy some of the query conditions. Various indexing structures [33,34] for spatiotemporal aggregate data are aR-tree, a3DR-tree, aRB-tree, aHRB-tree and a3DRB-tree. The original R-tree is extended in aggregate R-tree (aR-Tree) [33,34]. Each intermediate node in R-tree is augmented with storage of aggregated data for sub-branch under the node. aR-tree is generalized to three dimensional space in aggregate 3DR-tree (a3DR-tree) [33,34]. Both temporal and spatial dimensions can be indexed simultaneously in a3DR-tree. The drawbacks of a3DR-tree are its huge size that impacts query performance and wasting storage area by storing MBR for each change in aggregate data. The problems in a3DR-tree are rectified in aggregate RB-tree (aRB-tree) [33,34] by using R-tree to index aggregate spatial data and creation of a B-tree to each entry of the R-tree for storing historical aggregate data. If finest granularity in spatial dimension is considered, the regions become volatile. An approach to handle volatile regions is the creation of a new R-tree for each change. The aggregate Historical RB-tree (aHRB-tree) [33,34] is a combination of aRB-trees and HR-trees. The valid period of a spatiotemporal object during its lifespan is indicated by each node in HR-tree. A node is duplicated in HR-tree for each change in its entries. This causes redundancy of data in aHRB-tree. The aggregate 3DRB-tree (a3DRB-tree)[34] eliminates the data redundancy problem in aHRB-tree by combining 3DR-trees and B-trees.

VII. Miscellaneous Indexing Structures

Other indexing structures are Po-tree [35,36] and AIRSTD [35]. Po-tree is used to index real-time spatiotemporal data. It uses Kd-tree to index spatial aspects and B+ - tree to index the temporal aspects of the data. AIRSTD is an acronym for "Approach for Indexing and Retrieving Spatio-Temporal Data". It is a combination of 'Change Temporal Index' (CTI) and R-tree. CTI is used for temporal data indexing and R-tree is used for indexing spatial aspects.

VIII. Conclusions

All the spatiotemporal data indexing structures for accessing spatiotemporal data are collected and reviewed in this paper. They are organized into different categories and their purpose, focus and basis on which they are built is described. Indexing structures suitable for timestamp queries, range queries, trajectory queries and aggregate queries are specified. Standard bench mark datasets for evaluation and comparision of various spatiotemporal indexing structures are need of the hour to establish future direction for further research.

References

[1]. Mohamed F. Mokbel, Thanaa M. Ghanem and Walid G. Aref " Spatio-temporal Access Methods", IEEE Data Engg, Bull., 26(2):40-49,2003.

[2]. Long-Van Nguyen-Dinh, Walid G. Arif and Mohamed F. Mokbel " Spatio-Temporal Access Methods: Part 2(2003-2010). IEEE Data Engineering, Bulletin:

[3]. Peter Kuba, " Data Structures for Spatial Data Mining", FIMU-RS-200-05, September 2001.

[4]. Yufei Tao and Dimitris Papadias, " The MV3R-Tree: A Spatio-Temporal Access Method for Timestamp and Interval Queris", Proceeding of the 27th VLDB conference, Roma, Italy, 2001.

[5]. Lei Chen, Vincent Oria and M. Tamar Ozsu, " A Multi-level Index Structure for Video Databases", Proceedings of Internaltional Workshop on Multimdia Transformation Systems, October 2002,USA, PP:28-37.

[6]. Yannis Theodoridis, Michael Vazirgiannis and Timos Sellis, " Spatio-Temporal Indexing for Large Multimedia Applications", Proceedings of the Third Internaltional Workshop on Multimdia Transformation Systems, June 1996, PP:441-448

[7]. D. Pfoser, C. S. Jensen, and Y. Theodoridis. Novel Approaches in Query Processing for Moving Object Trajectories. In *Proc. of the Intl. Conf. on Very Large Data Bases, VLDB*, pages 395–406, Sept. 2000

[8]. P. Zhou, D. Zhang, B. Salzberg, G. Cooperman, and G. Kollios. Close pair queries in moving object databases. In *GIS*, pages 2–11, 2005.

[9]. E. Frentzos. Indexing objects moving on fixed networks. In *SSTD*, pages 289–305, 2003.

[10]. V. T. De Almeida and R. H. G¨uting. Indexing the trajectories of moving objects in networks. *Geoinformatica*, 9(1):33–60, 2005.

[11]. Y. Tao and D. Papadias. Efficient Historical R-trees. In *Proc. of the Intl. Conf. on Scientific and Statistical Database Management, SSDBM*, pages 223–232, July 2001.

[12]. Marios Hadjieleftheriou, George Kollios, Vassilis J. Tsotras and Dimitrios Gunopulos, " Indexing Spatio-temporal Archives", The VLDB Journal, Vol -….

[13]. Marios Hadjieleftheriou, George Kollios, Dimitrios Gunopulos and Vassilis J. Tsotras, "Efficient Indexing of Spatiotemporal Objects", Proceedings of the 8th International Conference on Extending Database Technology (EDBT), Springer-Verilog London, UK 2002, PP:251-268.

[14]. J. Ni and C. V. Ravishankar. PA-tree: A parametric indexing scheme for spatio-temporal trajectories. In *SSTD*, 2005.

[15]. T. T. T. Le and B. G. Nickerson. Efficient search of moving objects on a planar graph. In *GIS*, pages 1–4, 2008.

[16]. D. Pfoser, C. S. Jensen, and Y. Theodoridis. Novel Approaches in Query Processing for Moving Object Trajectories. In *Proc. of the Intl. Conf. on Very Large Data Bases, VLDB, pages 395-406, Sept. 2000*

[17]. L. Wang, Y. Zheng, X. Xie, and W.-Y. Ma. A flexible spatio-temporal indexing scheme for large-scale GPS track retrieval. In *MDM*, pages 1–8, 2008.

[18]. C. Jensen, H. Lu, and B. Yang. Indexing the trajectories of moving objects in symbolic indoor space. In *SSTD*, 2009.

[19]. Shengnan Ke, Jun Gong, Songnian, Qingzhu, Xintao Liu and Yeting Zhang, "A Hybrid Spatiotemporal Indexing Method for Tracjectory Databases", Sensors, July 2014.

[20]. X. Xiong, M. Mokbel, and W. G. Aref. LUGrid: Update-tolerant grid-based indexing for moving objects. In *MDM*, page 13, 2006.

[21]. Y. N. Silva, X. Xiong, and W. G. Aref. The RUM-tree: Supporting frequent updates in R-trees using memos. *VLDB J.*, 18(3):719–738, 2009.

[22]. K.-S. Kim, S.-W. Kim, T.-W. Kim, and K.-J. Li. Fast indexing and updating method for moving objects on road networks. *Web Information Systems Engineering Workshops*, 0:34–42, 2003.

[23]. J. Tayeb, O¨ . Ulusoy, and O.Wolfson. A Quadtree-BasedDynamic Attribute IndexingMethod. *The Computer Journal*, 41(3):185–200, 1998.

[24]. J. Dittrich, L. Blunschi, and M. A. Vaz Salles. Indexing moving objects using short-lived throwaway indexes. In *SSTD*, pages 189–207, 2009.

[25]. G. Kollios, D. Gunopulos, and V. J. Tsotras. On Indexing Mobile Objects. In *Proc. of the ACM Symp. on Principles of Database Systems, PODS*, pages 261–272, June 1999.

[26]. . Porkaew, I. Lazaridis, and S. Mehrotra. Querying Mobile Objects in Spatio-Temporal Databases. In *Proc. of the Intl. Symp. on Advances in Spatial and Temporal Databases, SSTD*, pages 59–78, Redondo Beach, CA, July 2001.

[27]. S. Saltenis, C. S. Jensen, S. T. Leutenegger, and M. A. Lopez. Indexing the Positions of Continuously Moving Objects. In *Proc. of the ACM Intl. Conf. on Management of Data, SIGMOD*, pages 331–342, May 2000.

[28]. M. Cai and P. Revesz. Parametric R-Tree: An Index Structure for Moving Objects. In *Proc. of the Intl. Conf. on Management of Data, COMAD*, Dec. 2000.

[29]. J.-D. Chen and X.-F. Meng. Indexing future trajectories of moving objects in a constrained network. *J. Comput. Sci. Technol.*, 22(2):245–251, 2007.

[30]. K Appathurai and S. Karthikeyan "A Survey on Spatio-Temporal Access Methods" , International Journal of Computer Applications (0975 – 8887), Vol-18, No-4, March 2011.

[31]. Z.-H. Liu, X.-L. Liu, J.-W. Ge, and H.-Y. Bae. Indexing large moving objects from past to future with PCFI+-index. In *COMAD*, pages 131–137, 2005.

[32]. D. Lin, C. Jensen, B. Ooi, and S. ˇ Saltenis. Efficient indexing of the historical, present, and future positions of moving objects. In *MDM*, pages 59–66, 2005.

[33]. Dimitris Papadias, Yufei Tao, Panos Kalnis and Jun Zhang, " Indexing Spatio-Temporal Data Warehoses",Proceedings of 18th International Conference on Data Engineering, San Jose, March 2002, PP:166-175.

[34]. Dimitris Papadias, Yufei Tao, Jun Zhang, Nikos Mamoulis, Qiongmao Shen and Jimeng Sun, " Indexing and Retrieving of Historical Aggregate Information about Moving Objects", Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 2002.

[35]. Hatem F. Halaoui, " AIRSTD: An Approach for Indexing and retrieving Spatio-Temporal Data", SITIS 2006.

[36]. Guillaume Noel, Sylvie Servigne and Robert Laurini, " Real-Time Spatiotemporal Data Indexing Structure", 7th AGILE Conference on Geographic Information Sciences, 29 april – 1 may, 2004, Greece.

SPIHT Compression Technique for DICOM Images using Wavelets

B. Janardhana Rao[1], O. Venkata Krishna[2]
[1]Department of ECE, CVR College of Engineering College, Hyderabad, India
Email: janardhan.bitra@gmail.com
[2]Department of EIE, CVR College of Engineering Hyderabad, India.
Email: venkatakrishna.odugu@gmail.com

***Abstract*—The image viewers do not consider specific requirements and scalability for the efficient encoding and decoding for easy transmission of images in many applications. This paper presents highly scalable Set Partitioning in Hierarchical Trees (SPIHT) algorithm for Digital Imaging and Communications in Medicine(DICOM) images. This algorithm is implemented using scalable line based Discrete Wavelet Transform(DWT) encoder and decoder. The performance metrics Mean Square Error (MSE) and Peak Signal to Noise Ratio(PSNR) are considered for achieving the high quality and scalable lossless image compression. This algorithm is also compatible with many heterogeneous networks.**

***Index terms*—MSE, PSNR, DWT, Image compression, SPIHT, DICOM images**

I. Introduction

The medical applications requires some standards for the memory storage purpose and for easy transmission [1]. The DICOM standard was developed to provide storage and some methods for easy transmission. The DICOM format has a header and a field in this header. The header always contains information about the patient, the image and imaging modality. The header field contains the 'Transfer Syntax Unique Identification'. This identification gives the type of Image compression technique which is used Compress the image data. The main disadvantage of the DICOM image is requires higher band width for transmission because usually the DICOM pictures and images are in uncompressed format and occupies more storage space. Presently, DICOM standard supports to different compression methods like lossless JPEG, lossy JPEG, JPEG-LS (lossless and near lossless), JPEG-2000 and run-length coding (RLE) [2]. In this paper, a progressive transmission coder, namely, set partitioning in hierarchical trees (SPIHT) method used to compress the DICOM images.

The aim of the work is to design an application for the easy accessing of DICOM images in any networks like UMTS, GPRS heterogeneous networks and so on. These DICOM images are compressed by systematically using of DWT and SPHIT techniques. The wavelet-based set partitioning in hierarchical trees (SPIHT) image coding algorithm generates a continuously scalable bit stream. This means that the images at various bit-rates and quality can be produced from single encoded bi stream, without any drop in compression. When a target rate or reconstruction quality has been reached then decoder simply stops decoding.

II. Discrete Wavelet Transform

The most important part which identifies many signals is the low-frequency content and the high-frequency content imparts details to the signal. The details and approximations are obtained after filtering in wavelet analysis. The details are high frequency components of the signal and low-scale whereas the approximations are the high-scale and low frequency components of the signal. When the original signal passes through two complementary filters it emerges two signals and it produce doubling of samples, then down-sampling is used to prevent this problem. The DWT coefficients re produced by the process on the right which includes down sampling. The decomposition is taken place on a signal, which is broken into many low resolution components. This entire successive approximated process is called wavelet decomposition is shown in figure 1.

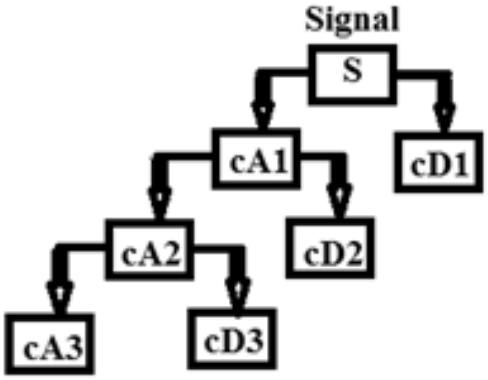


Figure 1: Decomposition in multi level

A) Wavelet Reconstruction

The Inverse Discrete Wavelet Transform (IDWT) is used for achieving the reconstruction of the image. The values are first up-sampled and then passed to the filters. The basis of the wavelet analysis is filtering and down-sampling, whereas up-sampling and filtering are the basis of the wavelet reconstruction process. The Up-sampling process means simply inserting the zeros between adjacent

samples then length of the signal component is increases. The original signal after up-sampling can be reconstructed from the coefficients of the approximations and details. Finally an approximated and detaild reconstructed signal with same length of the original signal is produced. These reconstructed details and approximations are known to be true constituents of the original signal. The down-sampling process is used for producing the details and approximations which are only half the length of the original signal. So, they cannot be directly combined to reproduce the signal [4]. Therefore, the approximations and the details are required to reconstruct before combining them. The reconstructed signal is represented as in figure 2.

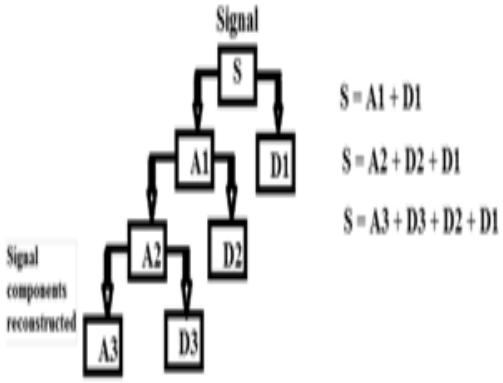


Figure 2: Signal components after reconstruction

The two-dimensional (2-D) wavelet transform can be extended from 1-D wavelet transform using separate wavelet filters. The 2-D wavelet transform of any signal can be obtained by using 1-D wavelet transform only. First by applying a 1-D wavelet transform to all the rows of the given signal, then same 1-D transform repeating on all columns of the signal [5]. The single level (K=1) 2-D wavelet transform of the original signal, with corresponding notation is shown in the Figure 3.

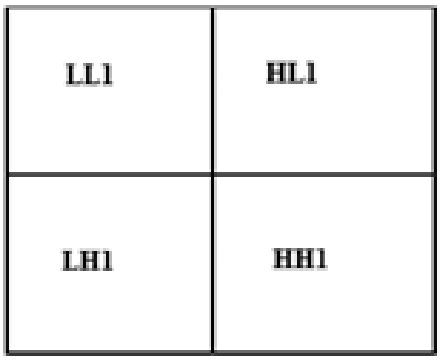


Figure 3: Single level sub band labeling scheme

The three level (K =3) wavelet expansion by repeating the same procedure is shown in Figure 4.

The highest level of the decomposition of the wavelet transform is represented by K in all discussions.

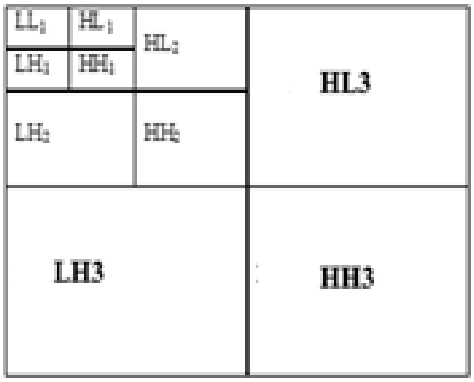


Figure 4: single level sub band labeling scheme

The 1-D version of decomposition is extended by the 2-D sub band decomposition. The 1-D decomposition method takes place two times; The first decomposition is taken place in horizontal direction and the second decomposition is in the orthogonal direction. The horizontal low pass sub bands(Li) further decomposed in the vertical direction, those are LLI and LHI sub bands. In the similar way high pass sub band (Hi) is further decomposed into HLi and HHi. After this process using the same method it is possible that the LLi sub band can be further decomposed into four sub bands and this produce multiple decomposition and multiple transform levels. The first level decomposition results the four sub bands are LL1, HL1, LH1 and HH1 which is shown in the figure 4. In the second level LL1 is further decomposed in to set of four sub bands LL2, HL2, LH2 and HH2. Next the sub band LL2 is used for the third level of transform or decomposition. In any band, the low-resolution sub band is LLi and sub bands LHi, HLi, HHi are the high-pass sub bands which represent the horizontal, vertical and diagonal sub band because they represents the residual information of the original image in the horizontal, vertical and diagonal directions respectively.

The four sub bands which are low-resolution, horizontal, vertical, and diagonal produced by applying the one-dimensional transform first along the rows and then along the columns to obtain a two-dimensional wavelet transform. At each level, the wavelet transform can be re-applied to the low-resolution sub band to next de-correlate the given image. The sub band conventions, image decomposition and the defining levels used in the AWIC algorithm which is illustrated in Figure 5. The final configuration contains a less low-resolution sub band. The original image data is referred by using phrase level 0. The original image data (level 0) is treated as a low-pass band when the user requests zero levels of transform and processing follows as its natural flow.

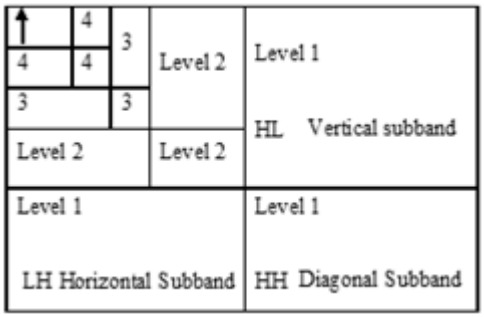


Figure 5: Image decomposition after wavelet transform

On each source image, the wavelet transformation is applied and the fusion decision map is generated. Using this map the fused wavelet coefficient map is generated from the original image wavelet coefficient. Then the fused image can be obtained from inverse wavelet transform. The fusion rules play a vital role in fusion process as explained in Figure 5.

III. Set Partitioning In Hierarchical Trees

The wavelet-based SPIHT image coding algorithm generates a continuously scalable bit stream. This means that the images at various bit-rates and quality can be produced from single encoded bi stream, without any drop in compression. When a target rate or reconstruction quality has been reached then decoder simply stops decoding. In the SPIHT algorithm, using hierarchical wavelet decomposition the image is first decomposed into a number of sub bands. Figure 6 shows the sub bands obtained for two-level decomposition. The spatial orientation trees are obtained by grouping the sub band coefficients, which is used for avoiding the correlation between the frequency bands. From the coefficients with highest magnitude the spatial orientation trees are coded bit plane by bit plane to the lowest pyramidal level. For further compression Arithmetic coding can also be used.

In general, the number of levels in the decomposition increases gives better compression results but the improvement becomes negligible beyond 5 levels. In practice the image dimensions limits the number of possible levels of decomposition because the wavelet decomposition can only be applied to images with even dimensions. The 5 level decomposition gives slight improved results using arithmetic coding only. To compress the binary file, a known prior information is required about the structure and properties of the image file and to exploit the abnormalities we have to assume the consistencies. The wavelet transformation produce the information about the image file then it is represented in a binary tree format. The tree contains a root and branches, where root with high probability of containing pixel magnitude level than that of the branches pixel magnitude. From this information, the algorithm is defined and named as the Set Partition in Hierarchical Tree (SPHT) algorithm.

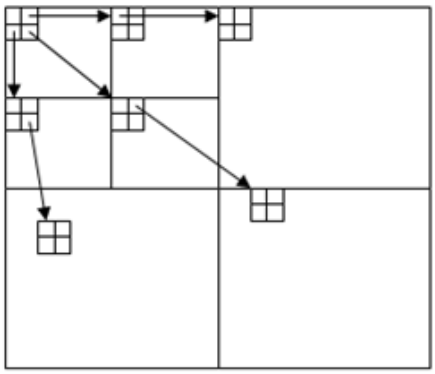


Figure 6: Spatial orientation tree by 2-level wavelet decomposition.

A) Set Partitioning Algorithm

The SPHT algorithm does not transmit the contents of the pixel values, or the pixel coordinates and the sets directly. It transmits only the decisions made in each and every step of the progression of the trees. Then these decisions define the structure of the image. These decisions gives the information of the pixel values , points, outcomes and part of the tree. Actually pixel coordinates are defined the part of the tree and pixel value defined by decision points and their outcomes. Finally, from these decisions the decoder also uses identical algorithm. This algorithm easily identifies the decisions and creates the identical sets as encode

The SPIHT part that gives the pixel values is the comparison of each pixel value. These pixel values designated to $2^n \leq c_{i,j} \leq 2^{n+1}$ with each pass of the algorithm by decreasing value of n. The decoder get the bit values each single value of n without any passing of pixel values. Then the magnitude of the compression can be controlled. For the sufficient number of n value, many information loops being passed with less error and for the small n value more pixel variation will be tolerated. Here $c_{i,j} \leq 2^n$ is a significant pixel value for given pass. Like this pixel values and coordinates are sorted by “significant’ and “insignificant “ tags. By using these significant and insignificant tags all the pixels are placed into partition sets. The problem with multiple times of traversing through all pixel values would take more time and inefficient.

Therefore to overcome the above problem the SPIHT algorithm is considered. In this algorithm, tree sort simulation is takes place only traverse into the tree as much as needed on each pass. This entire work is done by wavelet transform only. This wavelet transform produces properties of the image. Which are required for the algorithm. This algorithm "tree" is constructed as having the root at the very upper left most pixel values and this again extending downwards into the image. In this tree, the each node having four off spring nodes(See figure 6). The SPIHT method is an important advance method for image compression. It is not an extension from the traditional methods.

In image coding field, the image-coding using the wavelet transform is popular method. Generally so many coding algorithms are using wavelet transforms. But, there are two only successful, one is “embedded zero tree

wavelet coding" by Shapiro and the "set partitioning in hierarchical trees (SPIHT) " by Said and Pearlman. These two methods are very efficient and are able to reduce the block artifact than current standard JPEG image compression [6].

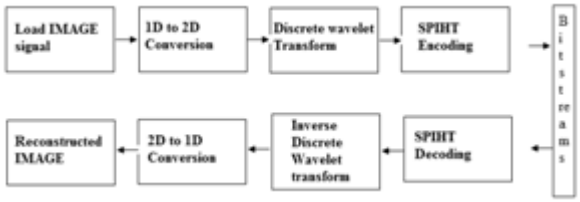


Figure 7: The block diagram for DWT-SPIHT encoding and decoding.

IV. Performance Evaluation

The important parameter for the image compression is the compression efficiency. Which is defined and estimated by the ratio of the original image size over the compressed data size.

The next parameter is complexity of an image compression algorithm. Generally the complexity is calculated with respect to encoding and decoding process in the algorithm i.e the number of data operations required to perform both encoding and decoding processes. But, sometimes it is expressed by the number of operations only.

In general, any encoding and decoding methods some data will be lost. Hence, the distortion measurement is also an important criterion to find out how much information has been lost. In image compression, the reconstructed image after compression is not exact replica of original image. This distortion is measured by the "mean square error(MSE)". The mean square error is given by:

$$MSE = \left(\frac{1}{N}\right)\sum_{i=0}^{N-1}\left(x_i^{'} - x_i\right)^2 \qquad \text{---- (1)}$$

where xi is the original signal value and $x_i^{'}$ is the reconstructed value . N is the total number of pixels. Mostly the MSE is measured in decibels for the lossy codec methods that is called "Peak signal to Noise Ratio". In this paper also the PSNR objective measurement for compression methods. The PSNR is given as,

$$PSNR = 10\log_{10}\frac{MAX^2}{\frac{1}{w*h}\sum_{i=1}^{w}\sum_{j=1}^{h}(o(i,j)-c(i,j))^2}$$

where h is the height of the image and w is the width of the image.MAX= 255 is maximum pixel value. c is compressed image date of original image o.

A) Error Metrics

Error metrics are used to compare the different image compression techniques. Two important error metrics 1) Mean Square-Error (MSE) and 2) Peak Signal-to-Noise Ratio (PSNR) are mostly used. The MSE gives the cumulative squared error between the original image and the compressed image and the PSNR shows the peak error[7].

The MSE and PSNR are given by,

$$MSE = \frac{1}{MN}\sum_{y=1}^{N}\sum_{x=1}^{M}[I(x,y) - I'(x,y)]^2$$

$$PSNR = 20 * \log_{10}(\frac{225}{sqrt(MSE)})$$

V. Simulation Results

The MATLAB code for Discrete wavelet transform is simulated. In this the original image is decomposed into high and low frequency signals. Next, the SPHIT algorithm encodes the resulting image of discrete wavelet transformation and sent to decompressing. The decompressing section consist of decoding and inverse discrete wavelet transform. Finally the approximated or reconstructed image to the original image is produced by the SPHIT.

The entire procedure, MATLAB code execution and the corresponding simulation results are shown in the following figures. The figure .8 represents the GUI (Graphical User Interface) window. In this window a DICOM image is browsed and the cho as in the figure .9.

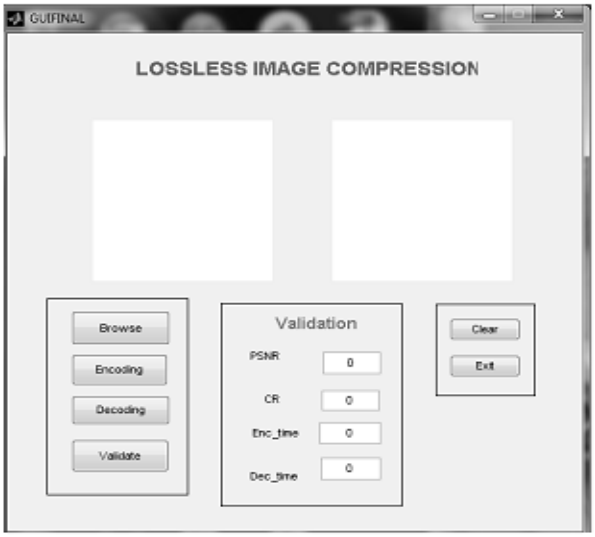


Figure 8: GUI Window after execution of code.

Figure 9: GUI window after browsing the image

In the GUI, by using the Encode tab the selected DICOM image is transformed into four sub-bands by Discrete Wavelet Transform. The resultant sub-bands are separated by low and high signals. Then, the transformed image is converted into bit-stream by encoding using SPIHT as shown infigure10

Figure 10: GUI window after encoding process

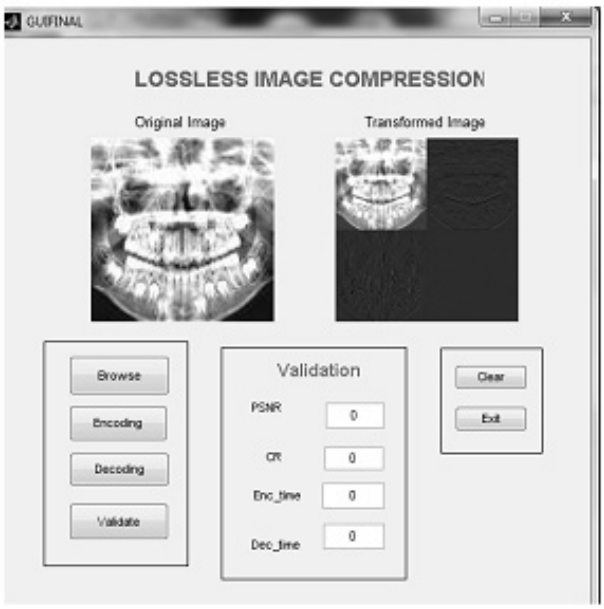


Figure 11: GUI window after encoding and ready for decoding process

At the receiving end the bit-stream is decoded and produce reconstructed image. Actually, the image is shown before decoding but the window is showing after decoding the image as shown in figure 12.

The figure. 13 represents the decompressed image or reconstructed image, which is almost the replica of the original image. This entire method provides lossless image with high compression ratio.

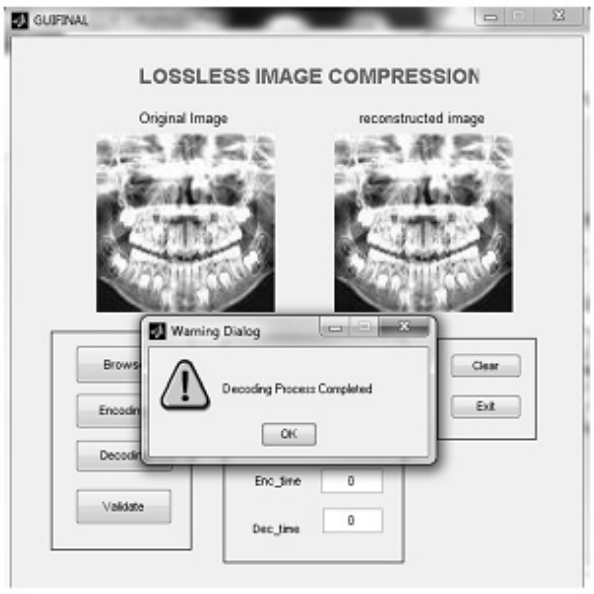


Figure 12: GUI window after decoding process

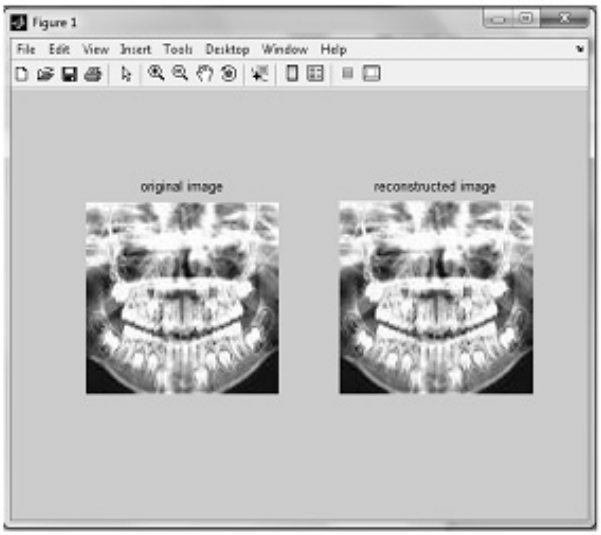


Figure 13: GUI window after reconstructed image

VI. CONCLUSIONS

In this paper, we presented a line based Discrete wavelet transform which applied to the encoder and decoder to perform image compression. In this algorithm the encoder and decoder consists of same memory and symmetrical with respect to complexity. Based on these parameters highly optimized SPIHT coding algorithm with less memory along with line based transform is presented. Comparatively this algorithm given good performance than many image compression coding methods. In this work the computational complexity and memory is reduced by the implementation of a less memory wavelet image coder. This wavelet image coder is optimized in terms of speed and memory. The further speed improvements can be achieved by using lifting steps or lattice factorization of the wavelet kernel.

REFERENCES

[1] DICOM Standard
http://medical.nema.org/dicom/2004

[2] DICOM Part 5: Data Structures and Encoding
http://medical.nema.org/dicom/2004/04_05PU.PDF

[3] J. Z. Wang, "Wavelets and imaging informatics: A review of the literature" J. Biomed. Inf., vol. 34, no. 2, pp. 129-141, Apr. 2001.

[4] Myint Myint Yi, Khin Sandar Linn, and Marlar Kyaw, "Implementation of Neural Network Based Electricity Load Forecasting," World Academy of
Science, Engineering and Technology 42 2008.

[5] Prarthana Shrestha "MIPS Augmented with Wavelet Transform: Perfor-mance Analysis", Computer Engineering Laboratory, Faculty of Information Technology and Systems, Delft University of Technology, The Netherlands, July 2002.

[6] J. Malý, P. Rajmic, "DWT-SPIHT IMAGE CODEC IMPLEMENTATION," Department of Telecommunications, Brno University of Technology, Brno, Czech Republic.

[7] Mallat, S.G. "A Thory of Multiresolution Signal Decomposition:The Wavelet Representation," IEEE Trans.PAMI, Vol.11, No.7, PP.674-693, July 1989.

Comparative Analysis of Higher Order Modulation Schemes using Simulink

P. Naga Malleswari
Department of ECE, CVR College of Engineering, Hyderabad, India
Email: pinjalamalleswari@gmail.com.

Abstract - **There is always a search for digital modulation technique that has efficient bandwidth and low bit error rate at a relatively low signal to noise power. Considering this requirement, in this paper we provide a general theoretical analysis of various M-ary PSK modulation for M = 2,4,8,16,and 32, M-ary QAM for M=4,16 and 64 and M-aryDPSK for M=2 and 4 using MATLAB Simulink taking BER as measure of performance when the system is subjected to Additive White Gaussian Channel. In this paper we take up bandwidth efficient linear digital modulation schemes and compare its performance based on their theoretical BER over AWGN and proves that increasing the value of M results in increase of BER with the increase in data rate.**

Index terms - **Bit error Rate (BER),Additive White Gaussian Noise (AWGN), M-ary PSK, M-ary QAM, M-ary DPSK and Signal-to Noise Ratio (SNR).**

I. INTRODUCTION

In digital communications industry the latest mathematical softwares can be used to increase the performance of digital system with several modulation techniques in digital like ASK (Amplitude Shift Keying), FSK(Frequency Shift Keying), PSK(Phase Shift Keying) PSK and QAM (Quadrature Amplitude Modulation) schemes. This paper contains both coherent and non coherent detection schemes. In later scheme, prior knowledge of the channel impulse response is unknown at the receiver. The phase of the received signal y is uniformly distributed between 0 and 2pi.This signaling scheme fails completely in the absence of noise. So this scheme is not a suitable method of detection especially in a fading environment. In coherent detection the receiver has sufficient knowledge about the channel impulse response. Pilot transmission technique is used to estimate the channel impulse response at the receiver. Equalization process gives the desired transmitted symbols ('x') from the receiver signal ('y') by using the following equation.

$$\hat{y} = \frac{y}{h} = \frac{hx + n}{h} = x + z \qquad (1)$$

Where z is an AWGN noise except for the scaling factor 1/h.

A. Bit Error Rate (BER)

Due to noise, distortion, interference, or bit synchronization errors are introduced in Digital transmission scheme. The Bit Error Rate (BER) is the number of bits in error divided by the total number of transferred bits during a studied time interval. BER is a unit less performance measure. The bit error rate Pb is given by

$$\text{Bit Error Rate}(P_b) = \frac{\text{Number of bits in error}}{\text{Total number of transfered bits}} \qquad (2)$$

Each modulation scheme performance can be measured by using probability of error with additive white gaussian noise. High data rate like 16-PSK can transmit four bits per symbol. To achieve a low BER, high power is not required. Modulation schemes are capable of delivering more bits per symbol are more immune to errors caused by noise and interference in the channel in communication system. Due to the increase of number of users, the mobile terminal is subjected to mobility. Thus the application of higher order modulations becomes a hot topic for many researches.

B. AWGN (Additive White Gaussian Noise) Channel

Generally noise is an unwanted signal presents in systems and the term additive indicates the noise is superimposed to the signal that tends to corrupt the signal where it will limit the rate of message transmission and limit the receiver ability to make correct symbol decisions [1].The received signal is the combination of transmitted and AWGN noise. The mathematical equation for Gaussian probability distribution function p(z) can be written as

$$p(z) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{z-a}{\sigma}\right)^2\right] \qquad (3)$$

Thus, AWGN is generated by thermal motion of electron in all dissipative electrical components like resistors, wires and so on. Mathematically, this noise is described by a zero-mean Gaussian random process where the random signal (z) is a sum of Gaussian noise random variable (n) and a dc signal (a) that is

$$z = a + n \qquad (4)$$

Thermal noise Power Spectral Density (PSD) is flat for all frequencies and is represented as

$$G_n(f) = \frac{N_0}{2} \qquad (5)$$

Equation (5) gives power spectral density with two sides equally distributed for both positive and negative frequencies [1]. When noise power has a uniform spectral density, then it is called as white noise. Since thermal noise is present in all communication systems with additive, white and Gaussian characteristics to model the noise in communication systems.

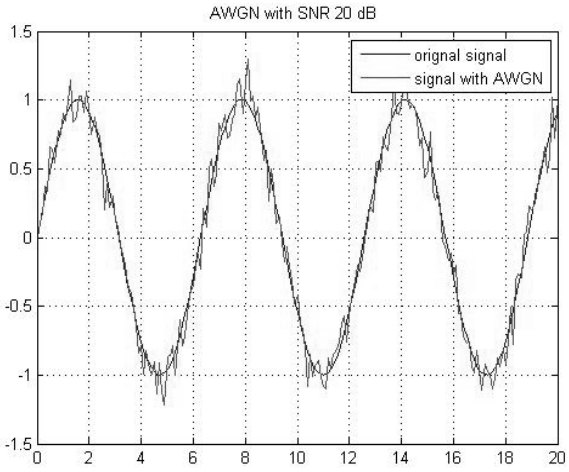


Fig.1: Example of AWGN signal

II. DIGITAL MODULATION TECHNIQUES

A. M-ARY PSK Modulation

In M-ary PSK BPSK ,QPSK ,8PSK,16PSK and 32PSK can be obtained by substituting M = 2,4,8,16,and 32.

BPSK is the simplest forms of PSK techniques which gives better performance compared to ASK and FSK. In this modulation, information is conveyed by varying the phase of constant amplitude carrier signal between two states which are separated by 180 degrees, i.e., the phase of carrier changes between 0 and 180degrees.

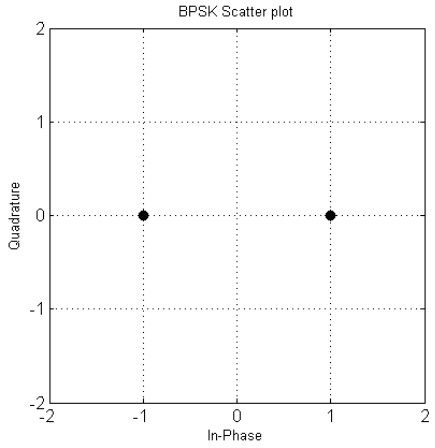


Fig.2: Constellation Diagram for BPSK

From Figure.2 symbols represents the constellation points of the signal. BPSK modulated signal is given by

$$S_1(t) = A\cos(2\pi f_c t)\ 0 \le t \le T \ \ for\ 1,$$

$$S_1(t) = -A\cos(2\pi f_c t)\ 0 \le t \le T \ \ for\ 0, \qquad (6)$$

these two signals have the same frequency and energy. BPSK bandwidth efficiency is more but low compared to other Mary PSK techniques. Low symbol rate of BPSK is not suitable for high data rate applications. This is mainly used for deep space telemetry. To avoid spectral spreading filtering can be employed.

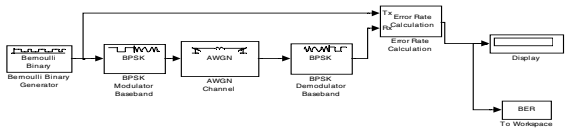


Fig. 3: BPSK Simulink Model

Figure.3 shows the Simulink based Binary Phase Shift Keying. In this model we used Bernoullis Binary data Generator,BPSK Modulator and Demodulator, Additive White Gaussian Noise Channel and Error Rate Calculation Blocks. All these block are collected from MATLAB Simulink Communication Block Set.

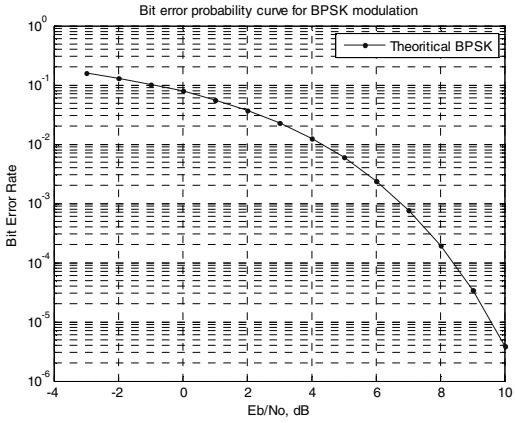


Fig.4: Bit error rate probability for BPSK over AWGN

Figure.4 shows the performance analysis of BPSK modulation over AWGN channel. It shows that the probability of error Vs Eb/No for BPSK.

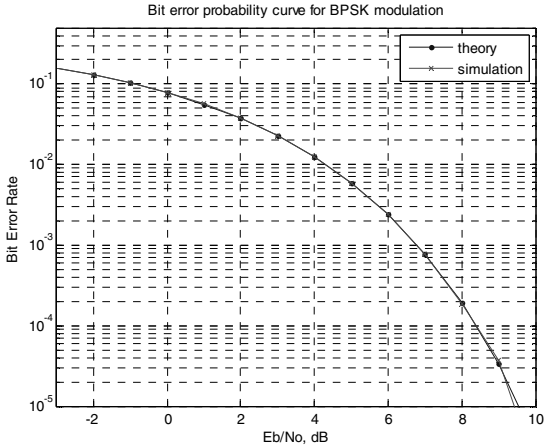


Fig.5: Theoritical vs Simulated analysis of BPSK over AWGN Channel using Simulink.

Figure.5 shows that the comparative analysis of theoretical and simulated Bit error rate for BPSK modulation over Additive White Gaussian Noise channel.

B. Quadrature Phase Shift Keying (QPSK)

QPSK is widely used in different applications as it does not suffer BER degradation. Like BPSK, information in QPSK is conveyed by shifting the phase of carrier.

The only difference is, the phase of carrier in this case is shifted between four different states separated by 90 degrees each. It has 4 states, two states are inphase component values and other two are quadrature component values.

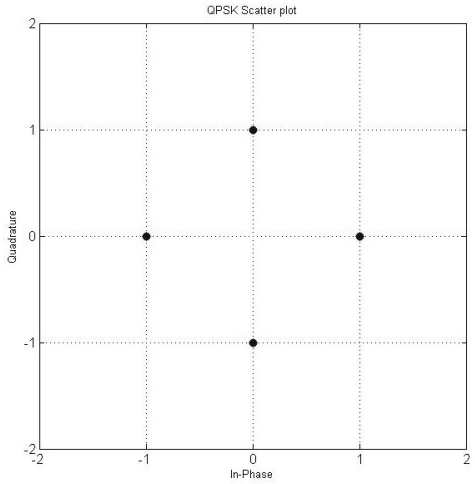


Fig.6: QPSK constellation diagram

Figure.6 shows that 4 constellation points with a phase difference of 90 degrees and equispaced around a circle. It encodes two bits per symbol which helps to reduce BER. QPSK provides same data rate as BPSK for half the bandwidth needed by BPSK. But the complexity of transmitters and receivers is the major problem in QPSK.

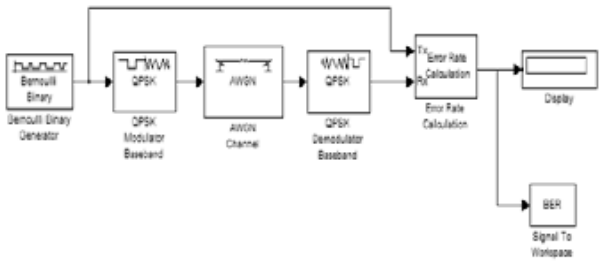


Fig.7: QPSK Simulink Model

Figure.7 shows simulink Model for Quadrature Phase Shift Keying. In general the MPSK signal is given by

$$S_n(t) = \sqrt{\frac{2E_s}{T_s}} \cos\left(2\pi f_c t + \frac{2\pi n}{M}\right) \qquad (7)$$

for QPSK technique M = 4 and n=1, 2, 3, 4.

The above equation with additional phase of $\frac{\pi}{4}$ can be written as

$$S_n(t) = \sqrt{\frac{2E_s}{T_s}} \cos\left(2\pi f_c t + \frac{2\pi n}{M} + \frac{\pi}{4}\right) \qquad (8)$$

$$S_n(t) = \sqrt{\frac{2E_s}{T_s}}\left(\cos(2\pi f_c t)\cos\left(\frac{2\pi n}{M}+\frac{\pi}{4}\right) - \sin(2\pi f_c t)\sin\left(\frac{2\pi n}{M}+\frac{\pi}{4}\right)\right) \qquad (9)$$

$\cos(2\pi f_c t)$ *and* $\sin(2\pi f_c t)$ in equation (9) represents two signals that are orthogonally separated. The In Phase and Quadrature components are given by

$$\varphi_1(t) = \sqrt{\frac{2E_s}{T_s}} \cos(2\pi f_c t) \rightarrow \textit{in phase component}$$

$$\varphi_1(t) = \sqrt{\frac{2E_s}{T_s}} \sin(2\pi f_c t) \rightarrow \textit{quadrature component}$$

So it consists of 4 constellation points given by

$$\left(\pm\sqrt{\frac{E_s}{2}} \pm\sqrt{\frac{E_s}{2}}\right) \qquad (10)$$

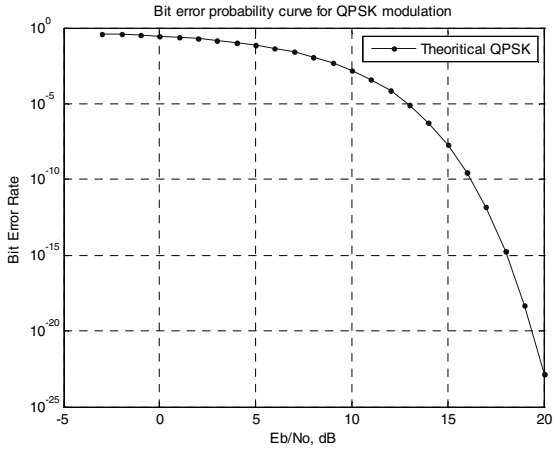


Fig. 8: Bit error rate probability for QPSK over AWGN

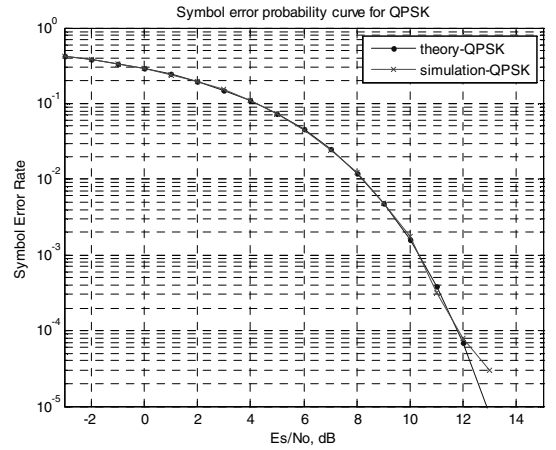


Fig.9: Comparative analysis of QPSK over AWGN Channel using MATLAB Simulink.

Figure.8 and 9 shows the theoritical and simulated bit error rate probability for QPSK over AWGN channel using MATLAB Simulink.

By substituting M=4 we obtain the value of P_s for QPSK. BER for BPSK is same as that of QPSK except channel bandwidth. It uses only half of the channel bandwidth and it transmits information at twice the bit rate of a BPSK system [5]. So it is mostly used in practice. The advantage of QPSK modulation is reduced bandwidth with increase in number of bits per symbol and disadvantage is that BER increases with increase in number of bits per symbol.

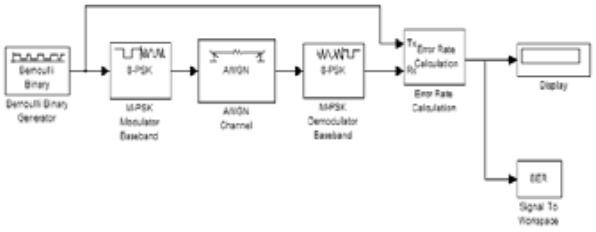


Fig.10: 8PSK Simulink Model

The 8PSK can be obtained by substituting M=3 and is also known as Octa Phase Shift Keying. The Bandwidth for OPSK is fb/3. Fig.11 shows the probability of error Vs Eb/No for OPSK.

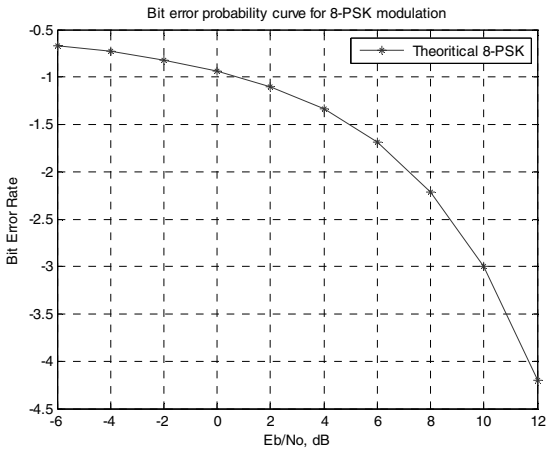


Fig.11: Bit error rate probability for 8-PSK over AWGN

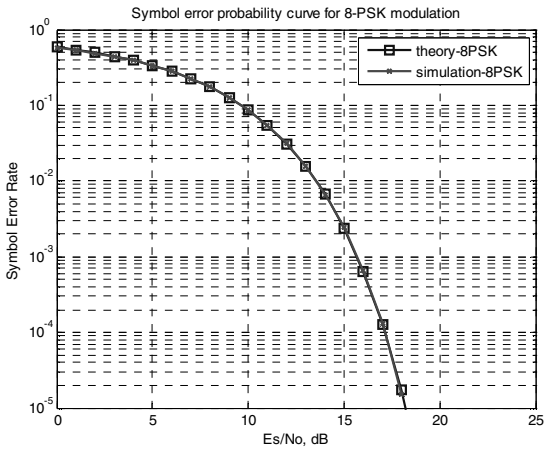


Fig.12: Comparative analysis of 8-PSK over AWGN Channel using MATLAB Simulink.

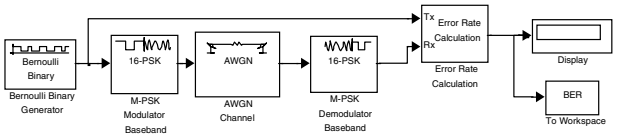


Fig: 13:16-PSK Simulink Model

Figure 13 shows the simulink model for 16-PSK and figure 14 and 15 shows the Theoritical and Simulated Bit Error Rates over AWGN Channel.

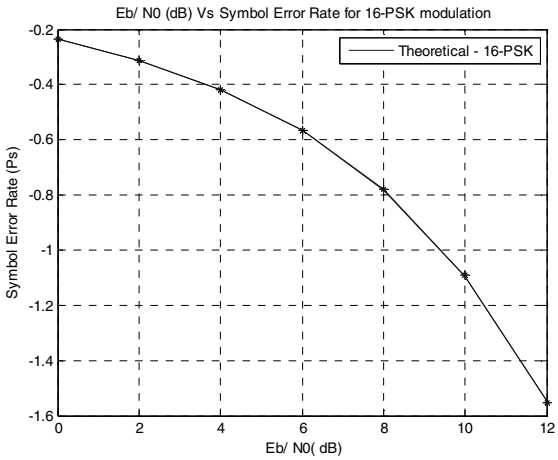


Fig.14: Bit error rate probability for 16-PSK over AWGN

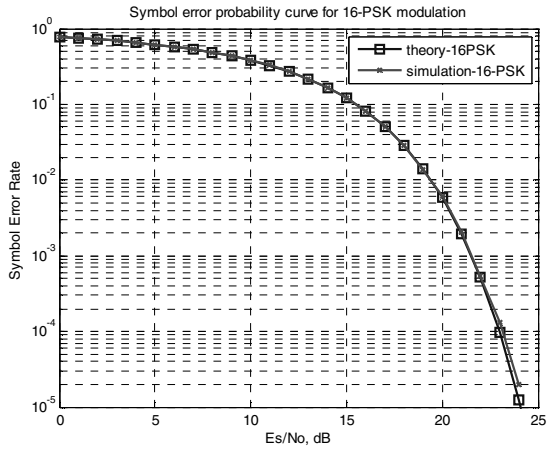


Fig.15: Comparative analysis of 16-PSK over AWGN Channel using MATLAB Simulink.

The 16-PSK is also known as Hexa phase shift keying. The Bandwidth for 16- PSK is fb/4. In this modulation the probability of errors and number of bits per symbols increases with the value of M. The constellation points come closer as the distance between them decreases.

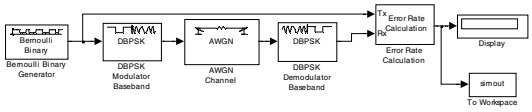


Fig.16:32-PSK Simulink Model

Figure.16 shows the Simulink Model for 32-PSK and figure 17 and 18 shows the comparative analysis of theoritical and simulated bit error rates.

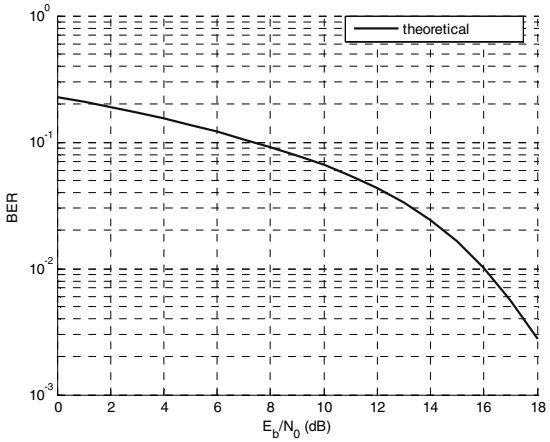


Fig.17: Bit error rate probability for 32PSK over AWGN

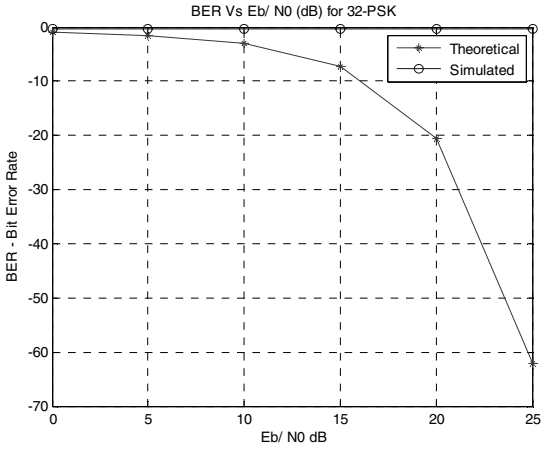


Fig.18: Comparative analysis of 32-PSK over AWGN Channel using MATLAB Simulink

DBPSK and DQPSK simulink model, theoritical and simulated bit error rate performances are shown in figure 19,20,21,22,23 and 24.

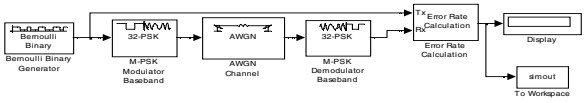


Fig.19: DBPSK Simulink Model

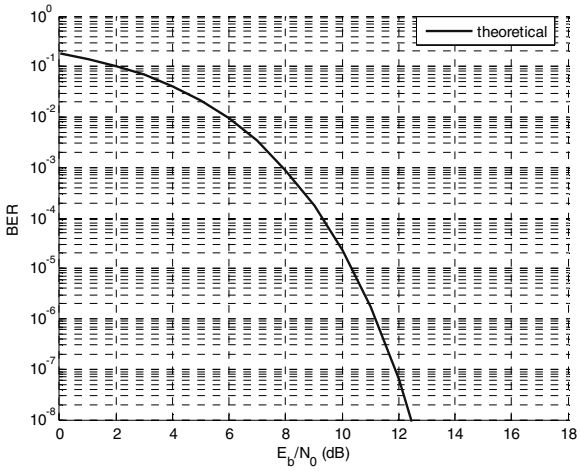


Fig.20: Bit error rate probability for DBPSK over AWGN

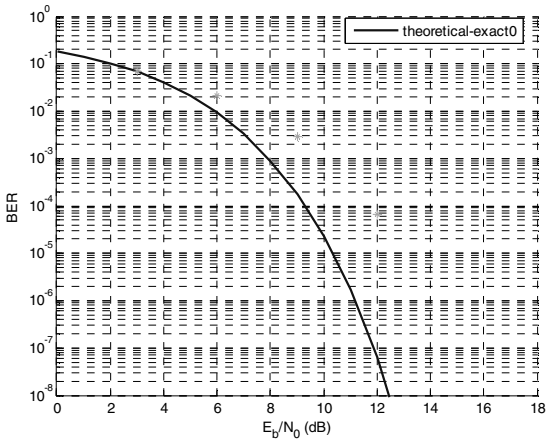


Fig. 21: Comparative analysis of DBPSK over AWGN Channel using MATLAB Simulink

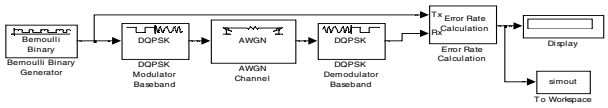


Fig.22: DQPSK Simulink Model

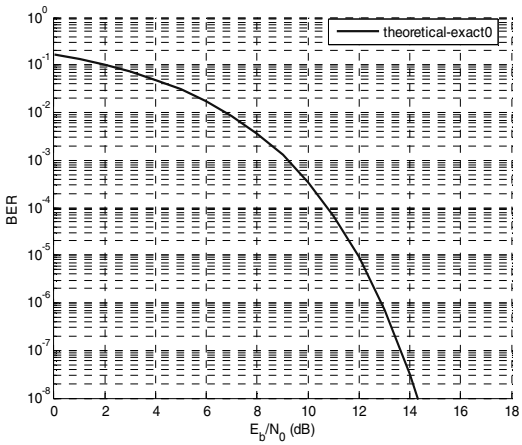


Fig.23: Bit error rate probability for DQPSK over AWGN

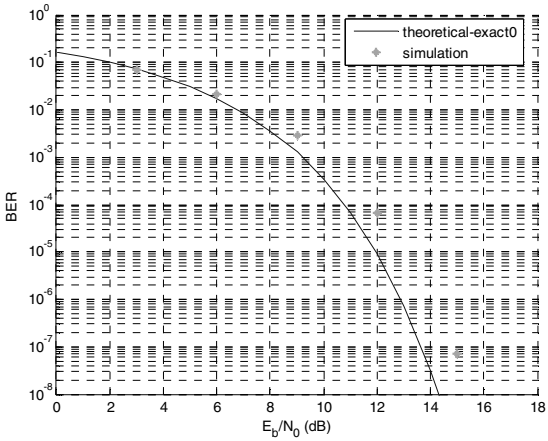


Fig.24: Comparative analysis of DQPSK over AWGN Channel using MATLAB Simulink

C. Quadrature Amplitude Modulation (QAM)

QAM transmits message by changing certain parameters of the carrier signal. In Quadrature Amplitude Modulation, amplitude is allowed to vary with phase,it is combination of ASK and PSK.But, implementation of this task is quite different. Unlike other techniques it Use s two quadrature carrier signals which are out of phase with each other by 90degrees. It transmits information by changing the amplitude of two carrier signals using ASK modulation technique. QAM can also be extended to M-QAM as 16QAM, 32QAM, 64QAM [2]. M-ary QAM modulation is more efficient than BPSK and QPSK.

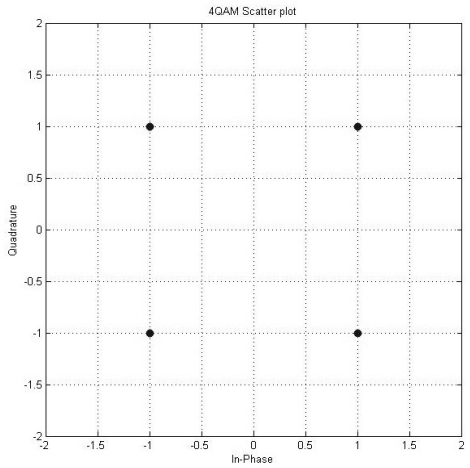


Fig.25: QAM Constellation

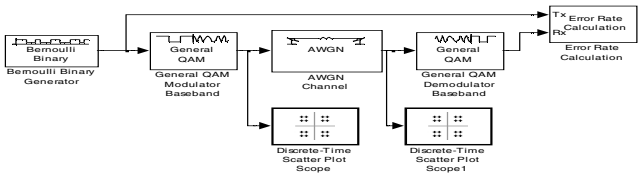


Fig.26: Simulink based M-ary QAM

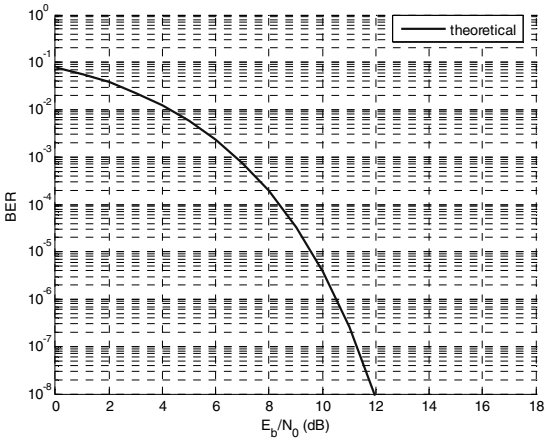


Fig.27: Bit error rate probability for 4QAM over AWGN

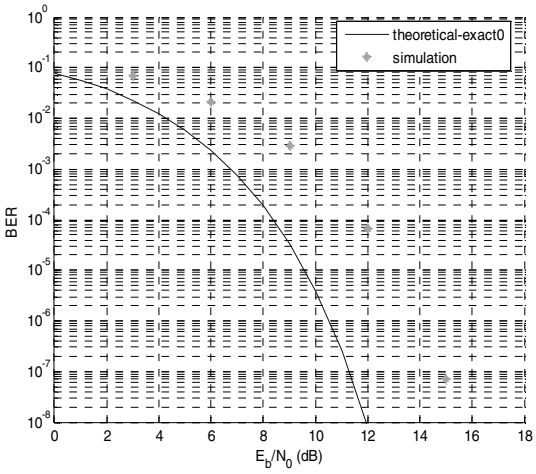


Fig.28: Comparative analysis of 4QAM over AWGN Channel using MATLAB Simulink

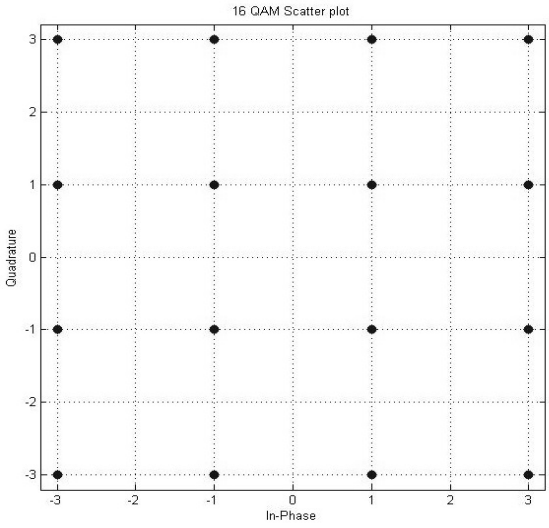


Fig.29:16-QAM Constellation

Similarly in other variations 16QAM four bits per symbol and symbol rate is one fourth of bit rate. 64QAM

six bits per symbol and its rate is one sixth of bit rate.Additive White Gaussian Noise (AWGN) Channel is widely used to analyse the modulation schemes in communication systems [3]. AWGN channel just adds a Gaussian noise to the signal passing through it without any loss of amplitude and phase distortion of frequency components. In this the received signal r(t) is sum of s(t) transmitted signal and White Gaussian noise n(t)

$$r(t) = s(t) + n(t) \qquad (11)$$

$n(t)$ is sample function of Additive White Noise with probability density function(Pdf) and power spectral density as $N(f)$.

$$N(f) = \frac{N_0}{2} \quad -\infty < f < \infty \qquad (12)$$

Here No is noise power density, an example of radio signal with AWGN noise signal is shown below with SNR of 20 dB.

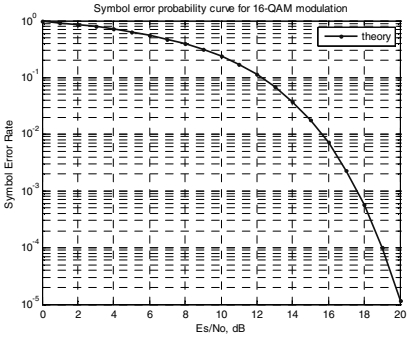


Fig.30: Bit error rate probability for 16QAM over AWGN

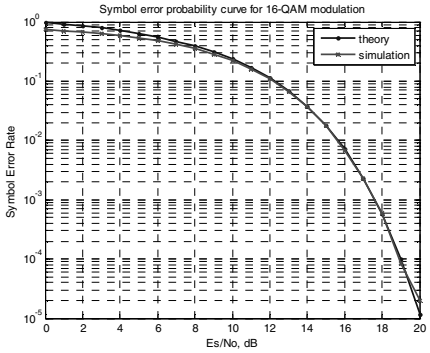


Fig.31:Comparative analysis of 16QAM over AWGN Channel using MATLAB Simulink

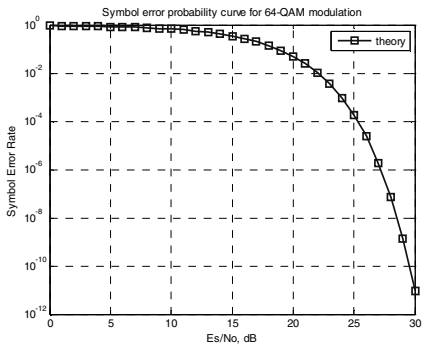


Fig.32:Bit error rate probability for 64QAM over AWGN

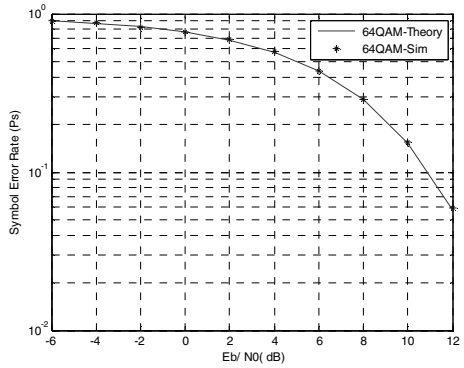


Fig.33: Comparative analysis of 64QAM over AWGN Channel using MATLAB Simulink

Theoretical BER for various linear modulations as shown in Table.1

Table 1: Theoretical BER over AWGN for various modulation techniques

Modulation	Detection Method	Bit Error Rate(P_b)
BPSK	Coherent	$0.5\ erfc\left(\sqrt{\frac{E_b}{N_0}}\right)$
QPSK	Coherent	$0.5\ erfc\left(\sqrt{\frac{E_b}{N_0}}\right)$
M-PSK	Coherent	$\frac{1}{m} erfc\left(\sqrt{\frac{mE_b}{N_0}} \sin(\frac{\pi}{M})\right)$
M-QAM(m=even)	Coherent	$\frac{2}{m}(1-\frac{1}{\sqrt{M}})erfc\left(\sqrt{\frac{3mE_b}{2(M-1)N}\right.}$
D-BPSK	Non- Coherent	$0.5e^{-\frac{E_b}{N_0}}$
D-QPSK	Non- Coherent	$Q_1(a, b) - 0.5\ I_0(ab)e^{-0.5(a^2+}$ $where\ a = \sqrt{\frac{2E_b}{N_0}(1-\frac{1}{\sqrt{2}}}$ $b = \sqrt{\frac{2E_b}{N_0}(1+\frac{1}{\sqrt{2}})}$ $Q_1(a,b) = MarcumQ - function$ $I_0(ab) = ModifiedBessel - function$

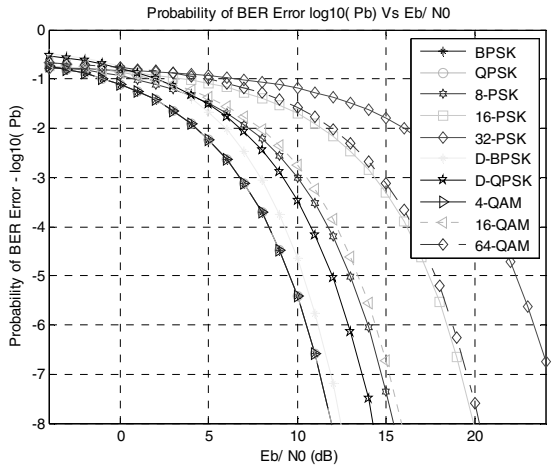


Fig.34: Comparative analysis of bit error rate probability for various modulation schemes over AWGN.

Fig.34 shows the comparison between BPSK, QPSK, 8-PSK, 16-PSK, 32PSK, DBPSK, DQPSK, 4QAM, 16QAM and 64 QAM modulation schemes. The probability of error Vs Eb/No for the above mentioned digital modulation techniques are obtained graphically by using MATLAB simulation.32PSK gives better results than previous modulation techniques[4].

III. SIMULATION RESULTS

The results of Bit Error Rate performance of M-ary PSK for M=2, 4,8,16 and 32 ,M-aryQAM for M=4,16and 64, DBPSK and DQPSK obtained using MATLAB SIMULINK Communication Toolbox. The overall comparative analysis of theoretical and simulated curves for BER vs Eb/ No (signal to noise ratio)for BPSK,QPSK,8PSK,16PSK,32PSK,DBPSK,DQPSK,4QAM,16QAM and 64QAM over AWGN channel are given in Fig.34.

IV. CONCLUSIONS

In this paper the mathematical analysis and simulations using Matlab shows that the Bit Error Rate for the M-ary PSK, DBPSK,DQPSK and QAM modulation techniques decrease monotonically with increasing values of SNR(Eb/N0). In QPSK system data transmission rate is twice the bit rate of a BPSK system . 8-PSK uses one third of the BW of BPSK and it transmits message at thrice the bit rate of a BPSK system that is clearly observed from the simulation curves. It is clearly observed that as the number of number of M increases, the error probability also increases over AWGN channel. Thus these modulations exhibit higher error-rates. Increasing the data rate will increase the SNR, however, increasing Rb (Bit rate in bits /second) will also cause more noise , since more bits are packed closer and sent through the channel. So, we cannot increase SNR by simply increasing Rb.

FUTURE SCOPE

The M-ary PSK can be implemented further by increasing M value by using more number of bits per symbol in both MATLAB and Simulink.

REFERENCES

[1] Harjot Kaur,Bindiya Jain, Amit Verma.”Comparative Performance Analysis of M-ary PSK Modulation Schemes using Simulink,” IJECT Vol. 2, Issue 3, Sept. 2011.

[2] Xiaolong Li, Indiana State University.”Simulink-based Simulation of Quadrature Amplitude Modulation (QAM) System”,Proceedings of The 2008 IAJC-IJME International Conference ISBN 978-1-60643-379-9.

[3] Sklar, B.“Digital Communications: Fundamentals and Applications”, Prentice-Hall, 2nd Edition, pp. 30-33,2001.

[4] MATLAB Help Documents, Communications Toolbox.2009.

[5] Ho, W. S., "Adaptive modulation (QPSK, QAM),"www.intel.com/netcomms/technologies/wimax/303788.pdf, accessed on December 30, 2007.

Automatic Reduction of Vehicular Speed in Speed Restricted Areas using Wireless Technology

Shakeel Ahmed[1] and Humaira Nishat[2]
[1,2] Department of ECE, CVR College of Engineering, Hyderabad, India
Email: shakeel_be@yahoo.com, huma_nisha@yahoo.com

***Abstract*—With the advanced technology being incorporated in automobiles, the acceleration in all the vehicles has been improved significantly. People using such vehicles have very little control on the upper speed limit of the vehicle while passing through small lanes, residential colonies as well as in speed restricted areas like school zones, banks, hospitals under construction areas etc. As a result the residents in such lanes had to resort to putting a number of speed breakers to avoid accidents. Erecting such speed breakers has many short term and long term effects both for the driver as well as the vehicle. Also, on main roads and highways, generally sign boards are kept which alerts the driver regarding speed limits at school zone, narrow road ahead, under construction area and so on so that the speed of the vehicle has to be reduced. But the driver of the vehicle either ignores the sign boards or the sign boards will not be visible due to the trees grown beside the road thereby covering the sign boards. Thus, there is a need to reduce the upper speed limit of vehicles. Also if this speed reduction is made possible without the intervention of the driver, the robustness of the action will be improved. Thus, this paper proposes the design of a wireless communication and control system that would control the upper speed limit of vehicles traveling on a certain road. The proposed system is simulated using Proteus 7 software and implemented using PIC Microcontroller (PIC16F877A), RF Module and other related components.**

***Index Terms*—alarm, driver, pic micro controller, roads, signboards, wireless communication system**

I. INTRODUCTION

Recent studies show that most of the road accidents are associated with excessive speed, carelessness of the drivers as well as the changes in the roadway i.e., frequent road construction work. Road accidents have become a big problem and a major cause of concern for the road traffic authorities, automobile industries and transport research groups. Thus, advanced driver assistance systems are incorporated in most of the vehicles today which allow the vehicle to communicate to the driver about the possibility of a collision through acoustic or visual signals produced by the vehicle. Future trends indicate that higher safety can be achieved by incorporating a large number of sensors both on the road as well as with the vehicle and through automatic driving controls. Cruise Control (CC) [3] is one of the driver assistance systems which can maintain a constant user preset speed. Adaptive Cruise Control (ACC) [1], [2], [4] which is addition to maintaining a constant user preset speed has the capability to maintain a safe distance from the preceding vehicles. But the major drawbacks of these systems are that the system is incapable of distinguishing between straight and curved parts of the road, the various speed limited restricted areas where speed has to be lowered to avoid accidents. A recently developed system known as Curve Warning Systems(CWS) uses a combination of Global Positioning System (GPS) and digital maps obtained from a Geographical Information System (GIS) which assess the threat levels of approaching a curve and alerts the driver quickly. Similarly Intelligent Speed Assistance (ISA) systems warns the driver about the inappropriate speed of vehicle using GPS and a digital road map containing information about the speed limits.

Despite all these systems, unexpected road circumstances like speed breakers, road work, road diversions, traffic congestion, etc., requires the attention of the driver while driving and have a control over the upper speed of the vehicle. Thus, some system has to be designed which would have a control or reduce the upper speed of the vehicle without the intervention of the driver in such speed restricted areas. This paper proposes the design of such a wireless communication system [9] which incorporates an RF transmitter at such speed limit zones/restricted areas and an RF receiver in the vehicles which receives the signals and in turn trigger a controller to reduce the upper speed of vehicles without the intervention of the driver.

The rest of the paper is organized as follows. Section II discusses the related work. Section III describes the proposed system with its working methodology and flowchart. Section IV discusses about the software used followed with the simulation results and section V gives the hardware results of the proposed system. The last section concludes the paper.

II. RELATED WORK

Different types of sign boards displaying the upper speed limit such as no over taking, humps, sharp curves, school zones, etc., are placed at the speed restricted areas to alert the driver by the road transport authority department as shown in fig. 2.1. It is up to the driver to follow them and take necessary action.

Fig. 2.1. Picture Showing Speed Limit Sign Board at Sharp Curves

In certain cases it also happens that the boards will be covered by trees that are growing beside the roads and will not be visible by the drivers as shown in fig. 2.2.

Fig. 2.2. Sign Board Covered With Trees

In order to avoid such situations one alternative is to incorporate CCS systems [3] also known as Speed Control or Autocruise or Tempomat in the vehicles. CCS is a system designed to automatically control the speed of the vehicle. It is a servomechanism system that takes over the throttle of the vehicle to maintain a steady speed asset by the driver. CCS is usually used for long drives across highways and sparsely populated roads which results in improved fuel efficiency. But when used on wet or snow and/or ice-covered roads, the vehicle may skid and results in driver losing control over the vehicle.

An alternative to CCS is Adaptive Cruise Control System (ACCS) [1], [2], [4] which is an improved cruise control. It has an automatic braking or dynamic set speed type controls. The automatic braking allows the vehicle to keep pace with the other vehicle and uses either a radar or laser setup. Some of such systems are also equipped with a feature called forward collision warning systems which warns the driver of the vehicle about the speed of the vehicles in front of it when it gets too close. Dynamic set speed type uses GPS positioning system for speed limit signs due to which the driver is relieved from the task of being careful while accelerating, deceleration and breaking in congested traffics. As the acceleration and breaking are done in a systematic way, the fuel efficiency of the vehicle is improved. But such systems enable the driver to become careless which may lead to severe accidents if the system fails.

In paper [5], authors have proposed a system to alert the driver about the speed limit at zones and to detect crashes automatically using GPS and GSM technology. The proposed work does not contribute for controlling the speed of vehicle without the intervention of the driver. Instead of controlling the vehicle speed automatically, the proposed system only alerts the driver about the speed limits and detects accidents.

In paper [6], authors have proposed a model to design an RF based speed control system for vehicles. But the system only alerts the driver by means of a buzzer and an LCD display regarding the zone in which the vehicle has entered. The system actually does not control the speed of the vehicle. By listening to the alarm, the driver himself should reduce the speed of the vehicle.

III. PROPOSED SYSTEM

The proposed system is divided into two sections-the transmitter section at the poles in each zone and the receiver section at the vehicle as shown in fig. 3.1. & fig. 3.2 respectively.

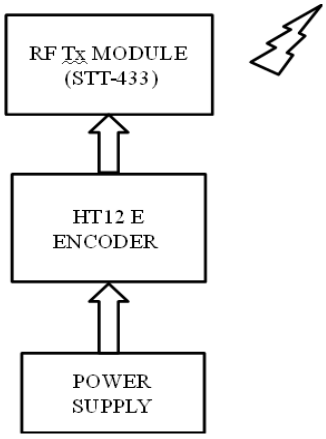


Fig. 3.1 Block diagram of the Proposed System- Transmitter Section

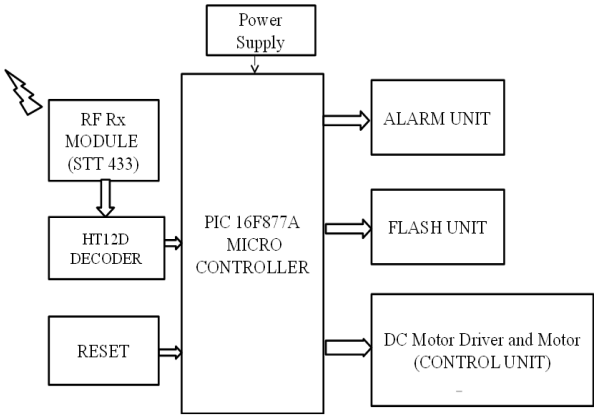


Fig. 3.2. Block Diagram of the Proposed System-Receiver Section

The transmitter section consists of a 433MHz ST-TX01-ASK Hybrid RF Transmitter Module (STT-433), GT12E Encoder and a Power Supply.

ST_TX01-ASK is an effective low cost, small size transmitter module with a frequency range of 315/433.92MHz, uses a supply voltage of 12V and generates an output power of 4 to 16 dBm. Fig. 3.3 shows the pin diagram of 433MHz ASK RF Transmitter.

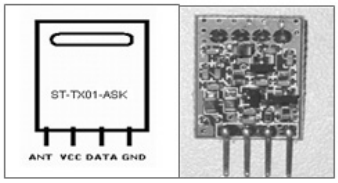


Fig. 3.3. Pin Diagram of 433 MHz RF Transmitter STT-433

HT12E 2^{12} series of encoders are a series of CMOS LSIs for remote control system applications. These encoders operate on a supply voltage of 0.3V to 13Vand are capable of encoding information which consists of N address bits and 12_N data bits. Each address/data input can be set to one of the two logic states. The programmed address/data are transmitted together with the header bits via an RF transmission medium. The 2^{12} series of encoders begin a 4-word transmission cycle upon receipt of a transmission enable (TE). This cycle repeats itself as long as the TE is held low. Once TE returns to high the encoder output completes its first cycle and stops. TE is enabled by applying a low signal to the TE pin as shown in fig. 3.4.

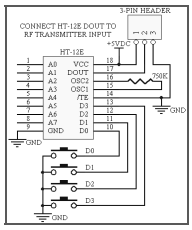


Fig. 3.4. Connections for HT12E Encoder

The receiver section comprises of a 433MHz ST-RX02-ASK Hybrid RF Receiver Module (STR-433), HT12D Decoder, PIC Microcontroller, Power Supply, DC Motor, Motor Driver (L293D), a buzzer and an LED.

ST-RX02-ASK is an effective low cost module available at frequencies 315/433.92MHz. STR-433 is ideal for short range remote control applications and requires no external RF components except for the antenna. It uses a super regenerative AM detector to demodulate the incoming AM carrier. Fig. 3.5 below shows the pin diagram of 433MHz RF Receiver STT 433.

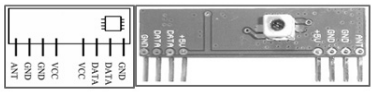


Fig. 3.5. Pin Diagram of 433 MHz RF Receiver STT-433

HT12D decoder belongs to 2 ^{12} series of decoder whose operating voltage is 2.4V to 12V. Is is a low power, high noise immunity CMOS technology capable of decoding 12 bits of information. These are paired with Holtek's 2^{12} series of encoders. For proper operation, a pair of encoder/decoder with the same number of addresses and data format should be chosen. The decoder receives serial addresses and data from a programmed 2^{12} series of encoders that are transmitted by a carrier using an RF transmission medium. Fig. 3.6 below shows the connections for HT12D Decoder.

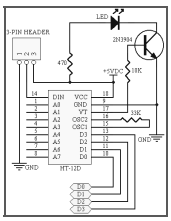


Fig. 3.6. Connections for HT12D Decoder

PIC Microcontroller (PIC16F877A) is an 8-bit microcontroller with 368 bytes of RAM, 256 bytes of EEPROM with a crystal oscillator operating at 20MHz and consists of 5 input/output ports viz. Port A, Port B, Port C, Port D and Port E. Each of the port can be used for a different function. Most of the ports can be used as either an input or an output port. Figure 3.7 below shows the architecture of PIC Microcontroller (PIC16F877A) and fig. 3.8 shows its pin diagram.

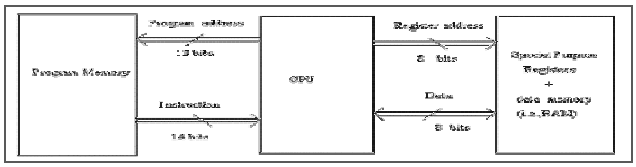


Fig. 3.7. Architecture of PIC Microcontroller (PIC16F877A)

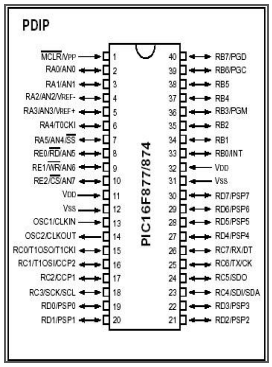


Fig. 3.8. Pin Diagram of PIC16F877A

Motor Driver (L293D): It is a dual H-bridge motor driver integrated circuit. Motor driver acts as a current amplifier by taking a low current control signal and providing a high current signal which is used to drive the motors. With L293D, two DC motors can be driven simultaneously both in forward and reverse direction. Logic 01 and 10 will rotate the motor in clockwise and anticlockwise directions respectively whereas logic 00 and 11 will stop it. Fig. 3.9a & 3.9b below shows the pin diagram and internal circuit diagram of L293D.

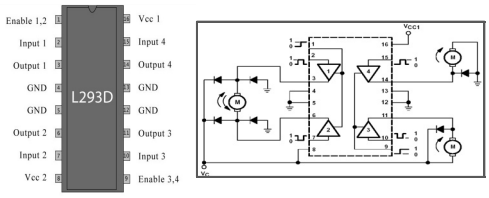


Fig. 3.9a. Pin Diagram of L293D; Fig. 3.9b. Internal Diagram

Buzzer: It is a sound source made up of piezoelectric sound component. By applying dc voltage between the electrodes of a piezoelectric diaphragm, mechanical distortions are caused due to the piezoelectric effect. To interface a buzzer the standard transistor interfacing circuit is used. Buzzer have a single tone and generates sound by using a battery as the power supply. Fig. 3.10 a & b below shows the picture of a buzzer and its transistor interfacing respectively.

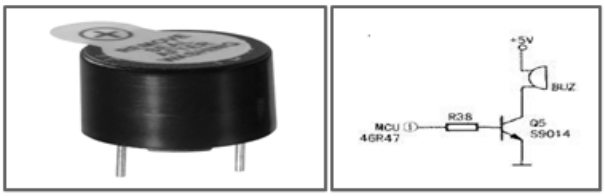


Fig. 3.10a. Picture of buzzer; Fig. 3.10b. Buzzer with transistor interfacing Circuit

LED: It is a semiconductor light source which is used as an indicator lamp in many devices.

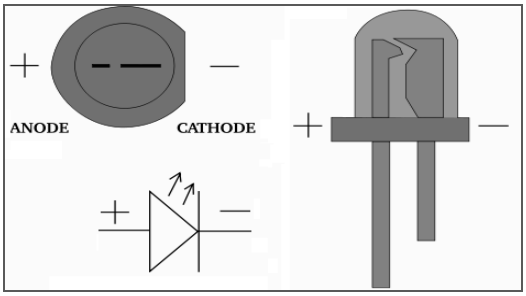


Fig. 3.11. Electrical Symbol & Polarities of LED

Working Methodology

The proposed system works in a simplex mode i.e., one way communication. When a vehicle enters a speed restricted lane such as a school zone, under construction area and so on the signal which is continuously radiated by the ASK RF transmitter section is received and decoded by the HT12D decoder and is applied as a triggering signal to the microcontroller. The flash unit or an LED connected to the circuit glows when valid data transmission occurs from transmitter to receiver. This indicator alerts the driver about the speed limit zone. Automatically without the intervention of the driver the upper speed limit of the vehicle is reduced. This actually happens when the microcontroller unit in turn triggers the control unit. In practice the control unit has to be the Electronic Control Unit (ECU) of the vehicle which controls the flow of fuel to reduce the speed of vehicle. But for the demonstration purpose, a vehicle is fabricated whose wheels are controlled by a dc motor driver and a dc motor. Thus, the control unit is replaced by a dc motor driver and a dc motor whose rotations are controlled and reduced on receiving a signal from the RF transmitter.

Thus, without the intervention of the driver the speed can be reduced. In case the speed control unit i.e., the control unit fails the driver is alerted by a hooting mechanism or the alarm unit (buzzer). The buzzer starts sounding so that the driver has to manually take the necessary action i.e., reduce the speed. The buzzer continues to sound unless the speed is reduced. Once the vehicle comes out of the speed restricted zone, due to the absence of any radiated signal the speed of the vehicle comes back to its normal value. Fig. 3.12 gives the flowchart of the proposed system.

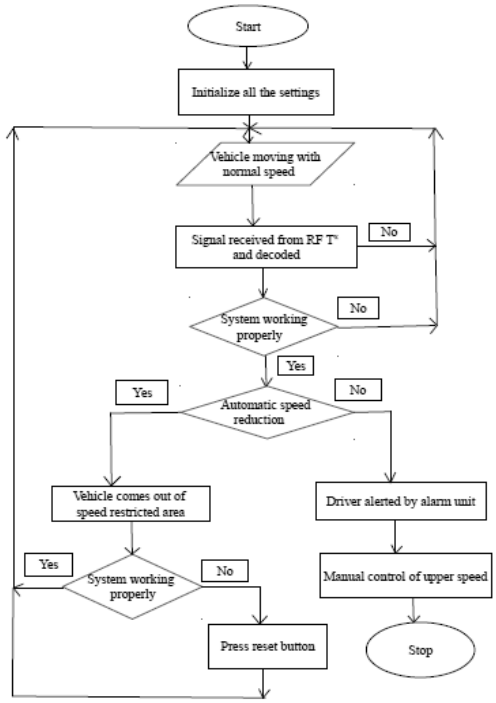


Fig. 3.12. Flowchart of the proposed system

IV. Simulation Results

The proposed system is simulated using the Proteus 7 software.

Simulation Results:

Fig. 4.1 shows a motor moving with the normal speed in the absence of any input data i.e., no signal being radiated by the transmitter. As the vehicle enters the speed restricted area then the input data is applied i.e., when the data key switch 1 to 4 is pressed then the speed of the motor is reduced accordingly and also the LED glows. This indicates that automatically the speed of the vehicle is reduced. The flash unit i.e., the LED alerts the driver about the speed restricted area and that he has no control over the upper speed limit of the vehicle.

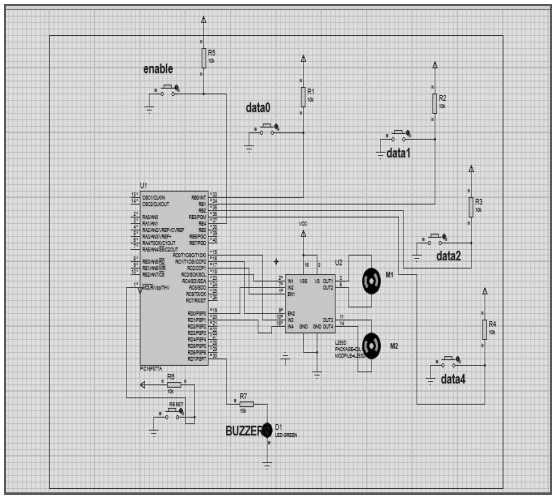


Fig. 4.1. Motor moving with normal speed when no data key is pressed

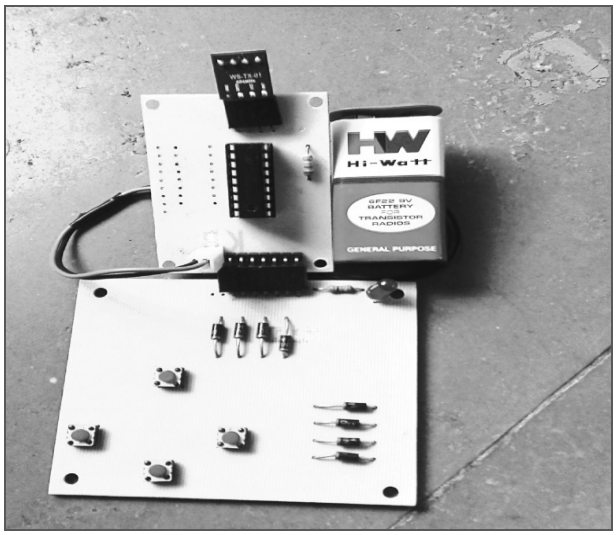


Fig. 5.1. Hardware Implemented Transmitter of the Proposed System

V. HARDWARE RESULTS

Fig. 5.1 shows the transmitter section of the proposed system. It comprises of a battery, four switches connected to the encoder which encodes the data and sends it through the RF transmitter.

Fig. 5.2 shows the receiver module which has to be placed and interfaced with the vehicle. The designed module reduces and limits the upper speed of the vehicle as it enters into a speed restricted area. When the switch at the transmitter side is pressed the LED at the transmitter glows and indicates that the data is being transmitted as shown in fig.5.3. Also the LED at the receiver glows indicating that the data is received and decoded by the receiver section and automatically the speed of the vehicle (here motor) is reduced.

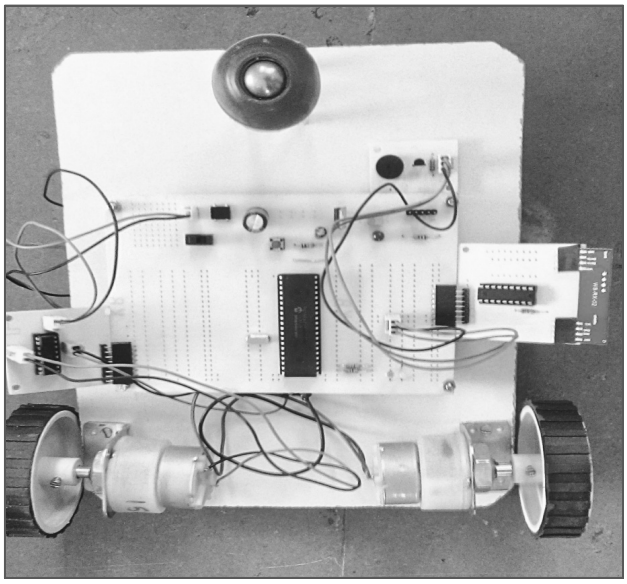


Fig.5.2. Receiver part placed in the vehicle

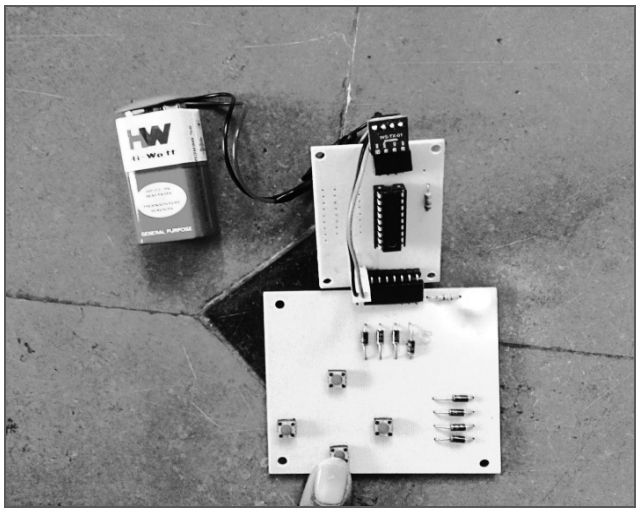


Fig. 5.3. LED Indicator Showing that One Key is Pressed

VI. CONCLUSIONS

The proposed system is realized using a dc motor and a motor driver using a PIC microcontroller instead of the ECU which controls the flow of fuel and reduces the speed of the vehicle. The proposed system can be implemented in real time and a working model can be developed which can be adapted by the automobile manufacturers and the road transport authorities. This system can prevent road accidents to a large extent as it avoids over speeding in the speed limit zones. It also takes care of wear and tear conditions of vehicles engine and driver health conditions as it is possible to avoid erecting speed breakers in lanes and speed restricted zones. Thus, the system can be easily implemented which ensures maximum safety not only to drivers but also to the passengers and pedestrians.

REFERENCES

[1] Ardalan Vahidi, Azim Eskandarian, “Research Advances in Intelligent Collision Avoidance and Adaptive Cruise Control”, IEEE Transactions on Intelligent Transportation Systems, vol.4, no.3, pp. 143-153, Sept. 2003.

[2] G.N.Bifulco, F.Simonelli, R.D Pace, “Experiments Towards Human Like Adaptive Cruise Control”, Proc. IEEE Intelligent Vehicles Symposium, Eindhoven, pp.919-928, 2008.

[3] F.A.Arvind Raj R., S.B. Sandhiya Kumar and T.C. Karthik S., “Cruise Control Operation from Zero to Preset Speed Simulation and Implementation”, International Journal of Information and Education Technology (IJIET), vol.1, no.1, pp. 9-14, April 2011.

[4] Shigeharu Miyata, Takashi Nakagami, Sei Kobayashi, Tomoji Izumi, Hisayoshi Naito, Akira Yanou, Hitomi Nakamura and Shin Takehara, “Improvement of Adaptive Cruise Control Performance”, EURASIP Journal on Advances in Signal Processing, vol. 2010, pp. 1-8, Dec. 2010.

[5] D.Narendra Singh, Ravi Teja ch.v, “Vehicle Speed Limit Alerting and Crash Detection System at Various Zones”,

International Journal of Latest Trends in Engineering and Technology (IJLTET), vol.2 Issue1, pp.108-113, Jan 2013.

[6] Ankita Mishra, Jyothi Solanki, Harshala Bakshi, Priyanka Saxena, Pranav Paranjpe, "Design of RF Based Speed Control System for Vehicles", International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), vol.1 Issue 8, Oct 2012.

[7] V.V.Sivaji, M.Sailaja, "Adaptive Cruise Control Systems for Vehicle Modeling Using Stop and Go Manoeuvres", International Journal of Engineering Research and Applications (IJERA), vol.3, Issue 4, pp.2453-2456, July-Aug 2013.

[8] M.Ben Swarup, M. Srinivasa Rao, "Safety Analysis of Adaptive Cruise Control System Using FMEA and FTA", International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), vol.4 Issue 6, pp.330-337, June 2014.

[9] V.Abinayaa, Anagha Jayan, "Case Study on Comparison of Wireless Technologies in Industrial Applications", International Journal of Scientific and Research Publications, vol.4 Issue 2, pp.1-4, Feb 2014.

[10] The "8051 Microcontroller and Embedded Systems" by Muhammad Ali Mazidi and Janice Gillispe Mazidi, Pearson Education.

[11] PIC Microcontroller Manual – Microchip

[12] PIC16F877A Data sheet

DFT based Channel Estimation in OFDM System

R. Prakash Kumar
Department of ECE, CVR College of Engineering, Hyderabad, India.
Email:Prakash.rachmagdu@gmail.com

Abstract- **Multi carrier modulation is Orthogonal Frequency Division Multiplexing (OFDM), dividing the large band spectrum into small (narrow) band spectrum. OFDM offers flexibility in adaptation to time varying channel condition by adopting the parameters at each subcarrier accurately. Due to the multipath effect the OFDM system suffers Inter Symbol Interference (ISI).ISI can be reduced by introducing the guard band (cyclic prefix).Time-varying channel estimation is an important activity in next generation wireless communication. Channel estimation is required mainly for the significant signal processing operation in multiple input multiple output MIMO- OFDM systems. In this project DFT – based channel estimation based on block type arrangement is implemented in MAT Lab. The bit error rate (BER) of DFT (Discrete Fourier Transform)based channel estimation is less than the BER of LS (Least Square) and MMSE (Minimum Mean Square Error) based channel estimation with same signal to noise ratio (SNR).**

Index terms: **OFDM, channel estimation, MMSE and LS.**

I. INTRODUCTION

During the past few years, there has been an explosion in wireless technology. This growth has opened a new dimension to future wireless communications whose ultimate goal is to provide universal personal and multimedia communication without regard to mobility or location with high data rates. To achieve these high data rates, the next generation during the past few years, there has been an explosion in wireless technology. This growth has opened a new dimension to future wireless communications whose ultimate goal is to provide universal personal and multimedia communication without regard to mobility or location with high data rate[3]s. To achieve these high data rates, the voice high quality, data, picture and video transmission should support in the next generation personal communication [2]. To transmit these services require data rate in terms of Mega bits per seconds (Mbps) Whenever data is transmitted high data rates, the response of the channel is spread over the many symbol periods, which leads to inter symbol interference (ISI)[4][10].To reduce the ISI usual technique is OFDM.In this technique the entire bandwidth is divided into many narrowband sub channels which are synchronous in transmission. All the sub channels or subcarriers are orthogonal to each other to eliminate the effect of delay spread. To reduce the co-channel interference in mobile environment and dispersive –fading limitations by adding turbo codes to OFDM system [4][5].

II. BLOCK DIAGRAM OF DFT BASED OFDM SYSTEM

The block diagram of the proposed system is shown in Fig1. At the receiver we are estimating the transmitted signal by adding the DFT block**.**

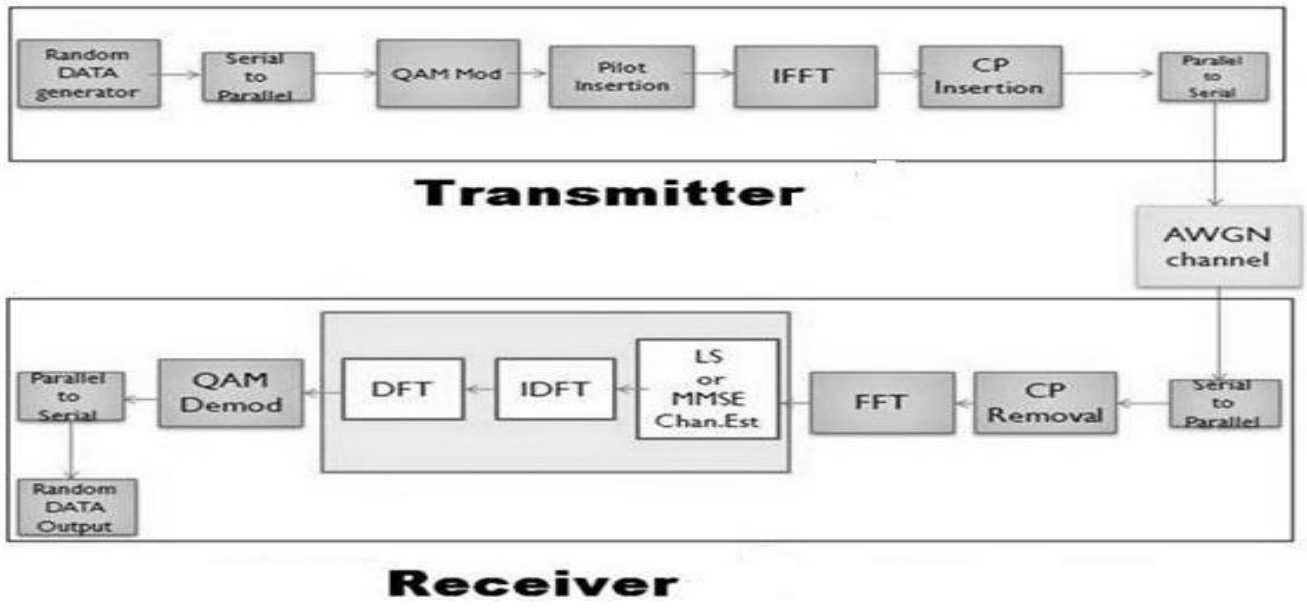


Fig1.Block diagram of DFT based OFDM system

III. DFT BASED CHANNEL ESTIMATION PERFORMANCE OF MSE

In the MIMO-OFDM context with Nt≤(M/CP) transmitter antennas and Nr receiver antennas, the (MSE) on the k -th subcarrier is equal to:

$$MSE(k) = \frac{\sum_{i=0}^{N_t} \sum_{i=0}^{N_r} E\left[||\widehat{H(k)} - H(k)||^2\right]}{N_t N_r}$$

Where H(k) and H(k) represent the estimated frequency channel response and the ideal one respectively.

For time and frequency selective MIMO ,MSE performance provided with FFT 2 ×2 MIMO system and its size is 1024.null symbol insertion is used for providing the Orthogonality between the subcarriers, for this subcarriers MSE is same and there is no border effect. By the impulse response from the first CP samples obtained the total power of the channel [7].

The performance of MSE and channel power is degraded due to null subcarrier are inserted on the edge of the spectrum [6], but number of subcarriers are increases the broader effect is observable.

LS and MMSE estimation is based on pilot frequency for comb-type.LS performance is not better than the LMS or MMSE complexity is more in MMSE. This will reduced by deriving the singular value decomposition with optimal low rank estimation [3][5]. Comb-type estimation depends on interpolation type such as second order, linear, spline. And low pass. Out of all these interpolations techniques second order interpolation has better performance than linear interpolation [3]. Low BER is achieved by time domain interpolation to linear interpolation [9].

IV. ALGORITHM FOR DFT-BASED CHANNEL ESTIMATION

The channel gain of the k^{th} subcarrier obtained from the LS channel estimation is denoted by $\hat{H}(k)$ taking IDFT of the LS estimation is

$$\{\hat{H}[k]\}_{k=0}^{N-1},$$

$$\begin{aligned} \text{IDFT}\{\hat{H}[k]\} &= h[n] + z[n] \\ &\triangleq \hat{h}[n], \end{aligned}$$

Where Z[n] denotes the noise component in the time domain. Ignoring the coefficients $\hat{h}[n]$ that contain the noise only, we can define the coefficients for the maximum channel delay L as

$$\hat{h}_{DFT}[n] = \begin{cases} h[n] + z[n], n = 0,1,2 \ldots L-1 \\ 0, \qquad\qquad otherwise \end{cases}$$

And transform the remaining L elements back to the frequency domain as follows

$$\hat{H}_{DFT}[k] = DFT\{\hat{h}_{DFT}[n]\}$$

The Algorithm of the performed task is given below.

Transmitter:

Step.1: Encode the message data by coding. In this paper we are using the Convolution coding.

Step.2: QAM is used for mapping of encoded data.

Step.3: for synchronize and estimating the channel some known bits are inserted to the modulated data called pilot symbols.

Step.4: Convert the time domain signal into the frequency domain by IFFT of step 3 data.

Step.5: To avoid the Inter symbol interference, insert the guard period between the output data of step.4.

Step.6: Transmit output data through the AWGN channel.

Receiver:

Step.7: received data having the noise, removed by using the filters. Guard period is removed and compute the FFT (i.e frequency domain into the time domain signal.)

Step.8: Channel estimate the filter co-effients[4] for received signal and other known information (such as modulation type and cyclic prefix length etc) apply the different kinds of estimating technique like LS, MMSE and DFT.

Step.9: Remove the channel effect from the output of step.7 with the help of the output found in step.8.

Step.10: Pilot removal, demodulation and decoding operation performed for the step 9 data.

Step.11: Calculate the Bit Error Rate (BER) for the step 10 data with channel estimation technique.

V. SIMULATION RESULTS

In this paper comb type pilot insertion scheme is implemented. We have discussed and simulated 3 types of channel estimation techniques[3][8]. Those are; LS-linear (Least Square Channel Estimation with Linear Interpolation), LS-Spline (Least Square Channel Estimation with Spline Interpolation) and MMSE (Minimum Mean Square Error) method [2].

These techniques are not that much efficient when compared to DFT based channel estimation. The Bit error rate (BER) and MSE (Mean Square Error) are improved in DFT- based channel estimation.

We have simulated the above mentioned techniques (LS-linear, LS-Spline and MMSE) in both scenarios. One with DFT- based Channel Estimation and another without DFT-based channel Estimation. In every method, the results of the method are compared with the true channel and MSE is calculated[10]. The number of symbol errors also calculated.

The Simulation Results are given below,

Case (i): SNR=10 dB

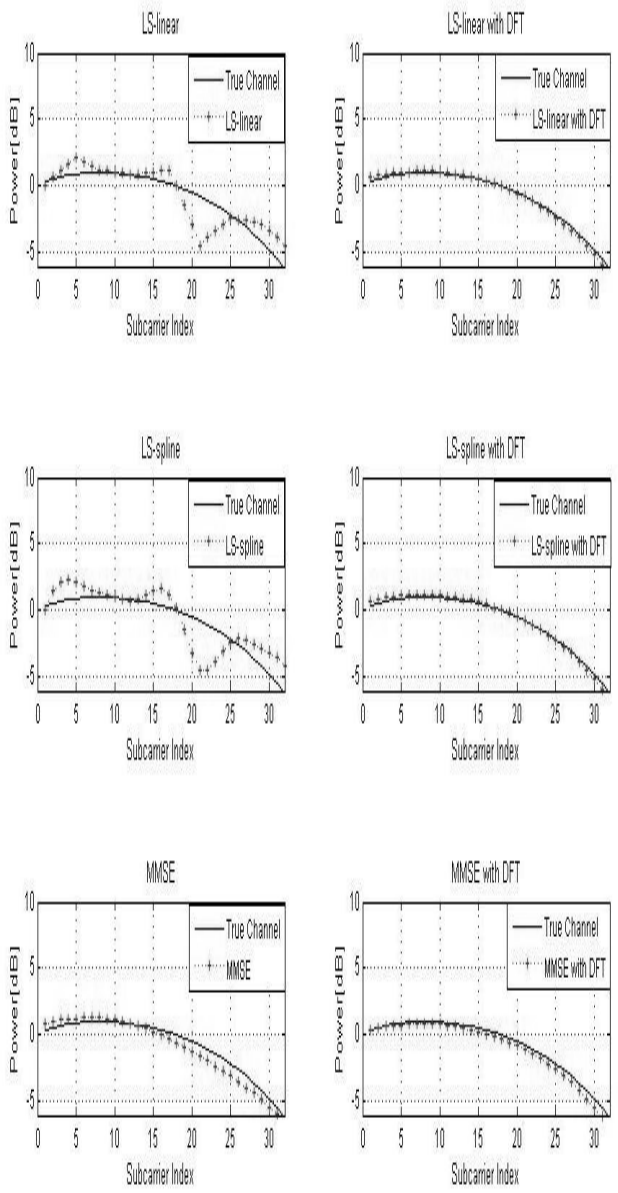


Fig.2: The Channel estimation of different methods with and without DFT based estimation for SNR=10dB

Fig.2 shows that Interpolation should be done to estimate the data. Interpolation can be done in two ways, one among them is Linear and the other one is spline interpolation[5]. In the Linear method, the samples are connected by using linear algebraic equation whereas in the Spline Interpolation the samples values are connected by using polynomial equations. For the lower SNR values(10dB in this case) , Linear and Spline Interpolation techniques fail to estimate the exact data[7].

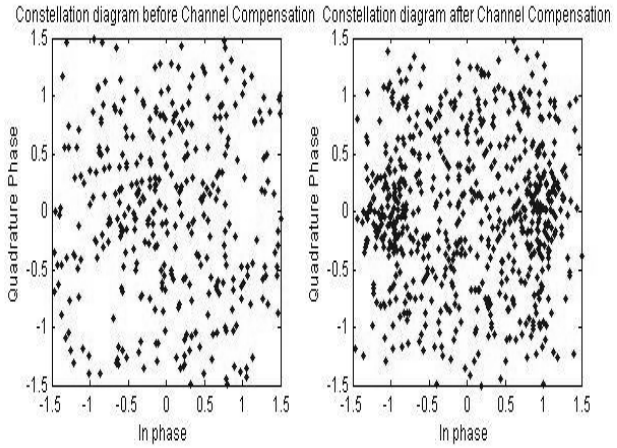


Fig.3: Comparison of Constellation diagrams before and after channel compensation for SNR=10dB

Constellation of QAM modulator is effected with AWGN (additive white Gaussian noise) component. Here we are comparing Constellation diagrams of QAM modulator before and after the channel compensation for SNR=10dB is shown in Fig.4.

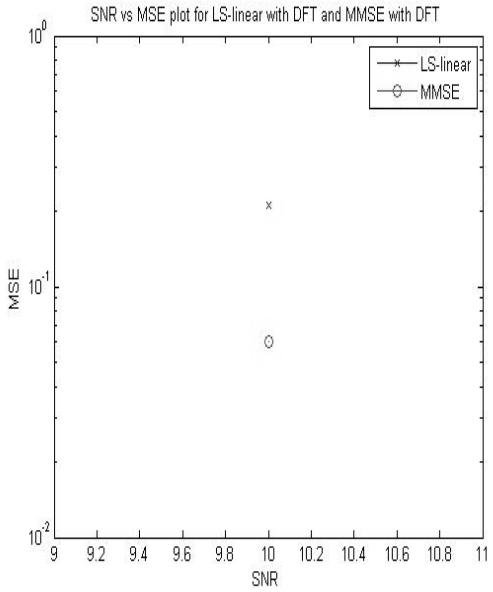


Fig.4: The Plot of SNR vs. MSE for SNR=10dB

SNR vs MSE graph is plotted for SNR=10dB.We can observe that the MSE for MMSE is less than the MSE of LS method.

Number of symbol errors = 1602

MSE of LS-linear =2.1246e-001 = 0.21246

MSE of LS-spline=2.6341e-001= 0.26341

MSE of MMSE Channel Estimation = 6.0107e-002= 0.060107

MSE of LS-linear Channel Estimation with DFT=3.6950e-002=0.036950

MSE of LS-spline Channel Estimation with DFT=3.7424e-002=0.037424MSE of MMSE Channel Estimation with DFT = 3.5421e-002=0.035421.

Case(ii) :SNR=30dB.

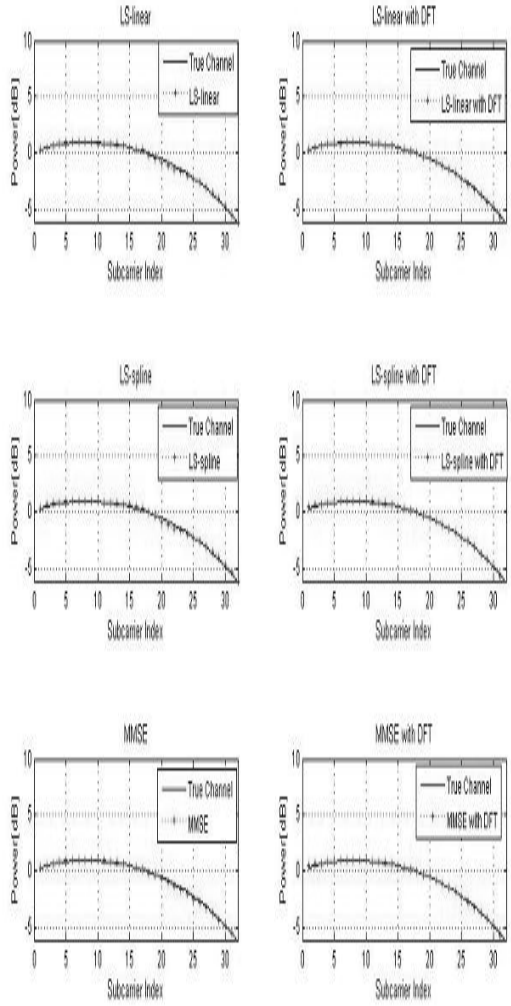


Fig.5: The Channel estimation of different methods with and without DFT based estimation for SNR=10dB

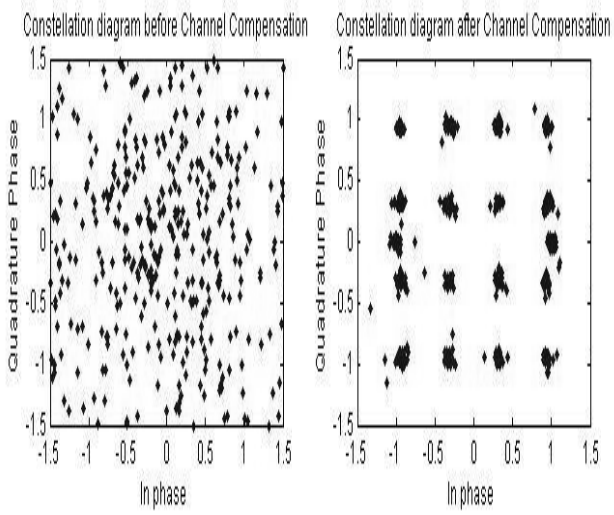


Fig.6: Comparison of Constellation diagrams before and after channel compensation for SNR=30dB

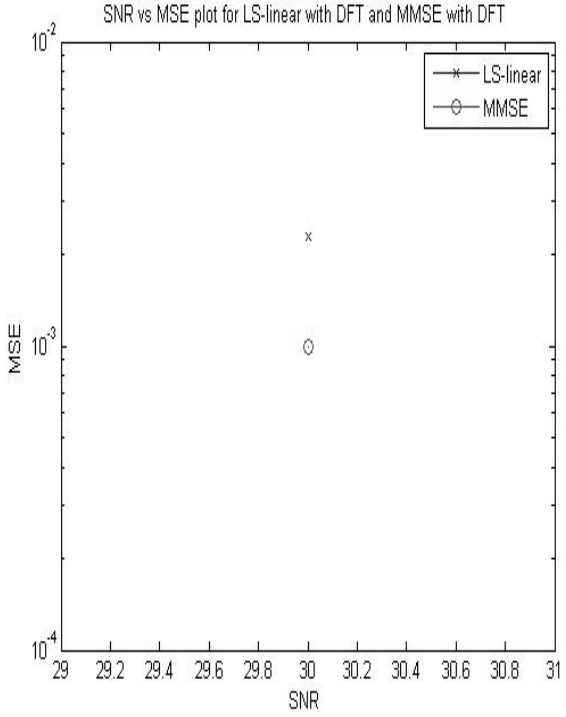


Fig.7 Plot of SNR vs. MSE for SNR=30dB

Number_of_symbol_errors = 14
MSE of LS-linear = 2.2877e-003=0.0022877
MSE of LS-spline = 2.7134e-003=0.0027134
MSE of MMSE Channel Estimation = 9.9295e-004 = 0.00099295
MSE of LS-linear Channel Estimation with DFT= 4.0627e-004 =0.00040627
MSE of LS-spline Channel Estimation with DFT=3.7935e-004 =0.00037935
MSE of MMSE Channel Estimation with DFT = 3.6774e-004 =0.00036774.

VI. CONCLUSIONS

Wireless communication transmits with high data rate and reliable from source to destination, this is achieved by OFDM technique. Major problem in OFDM system is fading effect[7][10]; this can be compensated at the receiver by channel estimation techniques. In this paper we are proposing the DFT channel estimation is better performance than the MMSE and LS techniques. The simulation is done for SNR of 30dB. The MSE for DFT based system dB is 0.00036774 and LS estimation is 0.00037935. The advantage of DFT based technique it does not require any prior channel information.

REFERENCES

[1] M. Sandell Edfors and J.-J. Van, "OFDM Channel Estimation By Singular Value Decomposition", IEEE Trans. Communication, vol. 46, no. 7, (1998) July, pp. 931-39.

[2] J.-J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson and P. O. Börjesson, "On Channel Estimation in OFDM Systems", IEEE 45th Vehicular Technology Conference, Chicago, vol. 2, **(1995)** July, pp. 815-819.

[3] Y. Zhao and A. Huang, "A Novel Channel Estimation Method for OFDM Mobile Communication Systems Based on Pilot Signals and Transform-Domain Processing", Vehicular Technology Conference, Phoenix, AZ, **(1997)** May, pp. 2089-93.

[4] A. Dowler, A. Doufexi, and A. Nix, "Performance Evaluation of Channel Estimation Techniques for a Mobile Fourth Generation Wide Area OFDM System", IEEE 56th Vehicular Technology Conference, Canada, vol. 4, **(2002)** September, pp. 2036-40.

[5] H. Minn and V. K. Bhargava, "An Investigation into Time-Domain Approach for OFDM Channel Estimation", IEEE Transactions on, Broadcast, vol. 46, no. 4, **(2000)** December, pp. 240-48.

[6] Y. Kang, K. Kim and H. Park, "Efficient DFT-based channel estimation for OFDM systems on multipath channels", IET. Communications, vol. 1, no. 2, **(2007)** April, pp. 197-202.

[7] Y. Lee, H. Shin, and H. Kim, "Channel Estimation Based on a Time-Domain Threshold for OFDM Systems", IEEE Transactions on, Broadcast, vol. 55, no. 3, **(2009)** September, pp. 656–663.

[8] J.-J. Van De Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, "OFDM Channel Estimation by Singular Value Decomposition", IEEE Trans. on Communications, vol. 46, no.7, (1998) July, pp. 931-939.

[9] Peter Hammarberg, Ove Edfors, "A comparison of DFT and SVD based channel estimation in MIMO OFDM systems", The 17th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'06), 2006.

[10] M.Belotserkovsky, "An equalizer initialization algorithm for OFDM receivers," Digest of Technical Papers International Conference on Consumer Electronics, 2002, pages 372–373, 20.

Cognitive Radio Networks- Spectrum Sensing

[1]M. Vinod Kumar Reddy, [2]K.Arun Kumar
[1,2]Department of ECE, CVR College of Engineering, Ibrahimpatnam, Hyderabad, India
Email: [1]vinodreddy.488@gmail.com, arun.katkoori@gmail.com

***Abstract*- Cognitive radio plays an important role in wireless communication. The spectrum has become a limited resource by ever increasing demand for wireless communications. For developing of new applications and services the existing spectrum allocation system creating a major problem. Cognitive radios are better solution to the problem of over-crowed and inefficient licensed spectrum. The sensing spectrum is the major key role in cognitive radio to find the free spectrum and favorable circumstances to use the unused frequency bands without causing intervention to the existing primary users(PUs).**

***Index terms* – Cognitive radio, spectrum sensing, energy detection, primary user, secondary user.**

I. INTRODUCTION

Wireless communications can be regarded as the most significant and important development in a modern society. Federal Communications Commission (FCC) found that spectrum access is more significant problem than physical scarcity of spectrum. With many developments in the area of wireless communication technologies already being employed Multimedia Broadcast and Multicast Services (MBMS) demand has tremendously increased and with the standardization of MBMS it has gained significant interest in the market[2]. Because It needs high bandwidth, storage capacity and some applications pose tight delay constraints, so it is necessary to optimize the utilization of spectrum is felt all the more.

Cognitive radio arises to be a solution to spectral crowding problem by introducing the opportunistic usage of frequency bands that are not highly occupied by licensed users since they cannot be utilized by users other than the license owners at the moment It is a system that senses its operational electromagnetic environment and can dynamically and automatically adjust its radio operating parameters to modify system operation, such as mitigate interference, enhance throughput facilitate interoperability, access secondary markets. Moreover, recent calculations by the Federal Communications Commission (FCC) show that the spectrum utilization in the 0-6 GHz band varies from 15 to 85% based on time, frequency and geographical area as shown in Fig. 1. These observations inspire the development of the cognitive radio (CR) and to modify the current static spectrum access policies accordingly in order to overcome the spectrum sacristy and underutilization problems.[4]

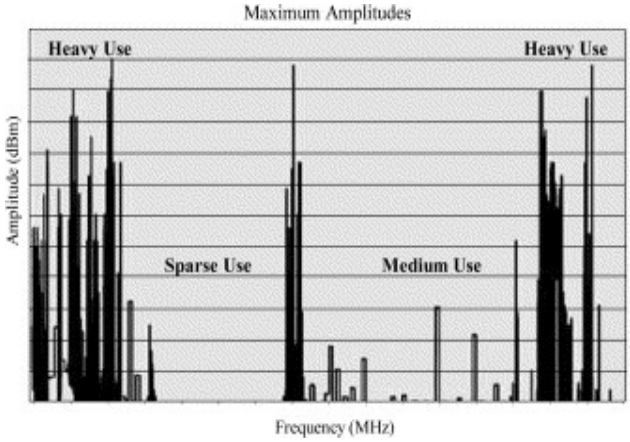


Figure 1: Spectrum Utilization

II. COGNITIVE RADIO

Cognitive radio is the technology which enables next generation communication networks, also known as dynamic spectrum access networks, to utilize the spectrum more efficiently in an opportunistic fashion without interfering with licensed users (primary users).

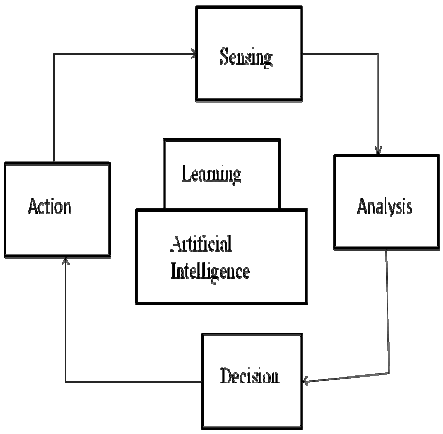


Figure 2: Cognitive Process

A basic cognitive process consists of following three basic tasks:

i. Sensing the spectrum
ii. Analyzing the Spectrum
iii. Decision Making of Spectrum

A. Sensing the Spectrum:

Spectrum sensing is the ability to measure, sense and be aware of the parameters related to the radio channel

characteristics, availability of spectrum and transmit power, interference and noise, radio's operating environment, user requirements and applications, available networks (infrastructures) and nodes, local policies and other operating restrictions. It is done across Frequency, Time, Geographical Space, Code and Phase.

B. Analyzing the Spectrum:

Spectrum Analysis is based on spectrum sensing which is analyzing the situation of several factors in the external and internal radio environment (such as radio frequency spectrum use by neighboring devices, user behavior and network state) and finding the optimal communication protocol and changing frequency or channel accordingly[6]. It is also known as channel estimation.

C. Decision Making of Spectrum:

Spectrum Decision Making calls for reconfiguration for the channel and protocol required for constantly adapting to mobile changing environments and adjustment of output power or even alteration of transmission parameters (such as modulation formats (e.g. low to high order QAM), variable symbol rates, different channel coding schemes) and characteristics by the Cognitive radio devices**.** CR should be able to use multiple antennas for interference nulling, capacity increase or range extension.

2.1 CHARACTERISTICS OF COGNITIVE RADIO

A. Cognitive capability: *B.*

Which is the ability to acquire the radio parameters from its surroundings.CR should be able to determine the frequency occupancy by identifying the spectrum holes (or spectrum opportunities). The spectrum hole is defined as the frequency bands which are allocated but not utilized in some location and at some times by the licensed system as given in Fig 3. Moreover, depending on the system, CR might have information about the modulation and coding as well as the relocation of the licensed system devices. *C.*

B. Reconfigurability:

Which is the ability to rapidly adapt the transmit parameters, i.e. operating frequency, modulation and coding, transmit power and communication technology, according to the radio environment in order to achieve the optimal performance

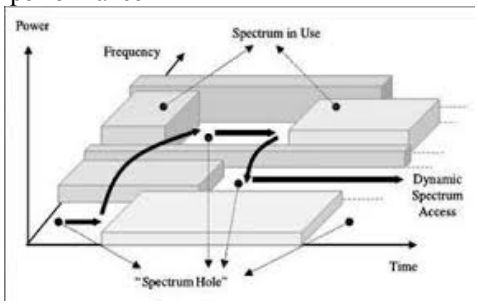


Figure 3: Spectrum Opportunities

2.2 FUNCTIONS OF COGNITIVE RADIO

The CR network has two main components: the primary network and the secondary one. The primary network, also referred to as licensed network, has a license to operate in a certain frequency band. It consists of primary users (PUs) with/without primary base stations (BSs). PUs is generally not equipped with any CR functions [8]. On the other side, the secondary network is able to access the licensed spectrum without affecting the primary network transmission. The secondary network is composed of secondary users (SUs) with/without secondary BS. Additionally, spectrum broker can be used to enable efficient and fair spectrum sharing between multiple secondary networks coexist in the same frequency band.

To support this type of spectrum sharing between the primary and cognitive networks, and to guarantee efficient usage of the resources in both networks, CR is required to perform the following four functions

A. A. Spectrum Sensing:

By this function, in order to identify the PUs activity the CR guides its radio environment Based on the sensing information, CR can determine the available spectrum holes that can be used for the CR transmission in a particular time, frequency, and location. Furthermore, the CR need to keep sensing the frequency spectrum during the CR transmission to avoid interfering with reappeared PUs.

B. Spectrum Decision:

This function analyzes the information from the spectrum sensing phase. The characteristics of the detected spectrum holes, the probability of the PU appearance, and the possible sensing errors should be considered before making the spectrum access decision [7]. Once the appropriate band is selected, the CR has to optimize the available system resources in order to achieve the required objective.

C. Spectrum Sharing:

This function chooses the appropriate MAC protocol to access the spectrum holes. By the MAC protocol, fair spectrum sharing between the different SUs can be guaranteed. Additionally, coordination between nodes can be achieved in order to avoid the collision with PUs as well as other SUs.

D. Spectrum Mobility:

Also called spectrum handover and by this function, CR is able to change the operating band in order to avoid a detected PU activity. Additionally, the CR can perform the spectrum handover in order to improve the secondary network performance by transmitting in another spectrum hole with better condition [6]. The protocol parameters at the different levels should be adapted according to the new operating band.

III. SPECTRUM SENSING

Spectrum sensing is universally recognized as main enabler for cognitive radio. The actual meaning of

spectrum sensing is to provide awareness of the radio environment gives indication on the availability of transmission opportunities, gives indication on who is occupying the channel and gives indication on the quality of channel itself.

3.1 BASICS-COGNITIVE RADIO -SPECTRUM SENSING

There are various of considerations in Frequency spectrum sensing occupancy. They are

A. Continuous Spectrum Sensing:

In the cognitive radio, we have to continuously sense the spectrum occupancy. Generally a cognitive radio employs the spectrum on a non-intervention basis to the primary user. Correspondingly, it is required for the cognitive radio system to continuously sense the spectrum in case the actual primary user returns.

B. Monitoring for alternative Empty Spectrum:

The cognitive radio system must have alternate spectrum available to which it can change should the right appear if the existing primary user returns to original spectrum which is being used.

C. Monitoring the type Of Transmission:

Whatever the type of transmission being received is required to sense in cognitive radios and cognitive radio system be able to find the type of transmission used by the actual primary user so that specious transmissions and interference are ignored as well as transmissions made by the cognitive radio system itself[4].

There are number of ways in which cognitive radios are able to perform frequency spectrum sensing. The ways in which cognitive radio frequency spectrum sensing can be performed in any one of the following methods:

i. Non-cooperative:

In this, frequency spectrum sensing occurs, when a CR system acts on its own. The cognitive radio system will configure its own according to the signals it can find the information with which it is preloaded.

ii. Cooperative:

In this, sensing will be initiated by a number of different radio systems within the cognitive radio network. Typically a central station will receive reports of signals from different radio stations in the network and adjust overall cognitive radio network to suit.

The problems of unnecessary interference is reduces in cognitive radio cooperation where a single cognitive radio cannot hear a primary user because of issues such as changing from primary user, but second primary user acts as a receiver may be able to hear both the primary and the signal from the cognitive radio system.

IV SPECTRUM SENSING METHODS

In this paper we discuss about some of the spectrum sensing techniques.
They are:

A. Energy Detection

It is most common way of spectrum sensing because of its low computational and implementation complexities [6]. This technique also known as radiometry and it is more generic as receivers do not require any knowledge on the Primary signal energy. It consists in extracting as a feature the incoming signal power, normalized by the noise power.

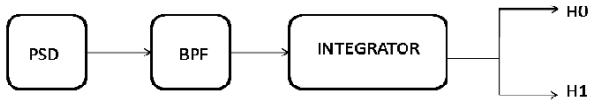


Figure 4: Block Diagram of Energy Detection

Where H0 Absence of user
H1 Presence of user

Block diagram for the energy detection technique is shown in Figure 4. In this, First the power spectral density of the signal is calculated and this is passed through band pass filter of the bandwidth W and is integrated over time interval. The output from integrator is then compared with the predefined threshold value. This comparison is used to find the existence or absence of the primary user and this value can set to be fixed or variable based on the channel conditions.

$y(k) = n(k)$…… H0
$y(k) = h * s(k) + n(k)$…… H1

y (k) is the sample to be analyzed at each instant k and n (k) is the noise of variance σ^2. Let y(k) be a sequence of received samples $k \in \{1, 2….N\}$ at the signal detector, then a decision rule can be stated as,

H0…… if $\varepsilon > v$ and H1…… if $\varepsilon < v$

Where $\varepsilon = E\,|y(k)|^2$ the estimated energy of the received signal and v is chosen to be the noise variance σ^2.Any how the energy detection has the following disadvantages

i) The Sensing time taken to achieve a given probability of detection may be more.
ii) The detection performance is subject to the uncertainty of noise power.
iii) The energy Detection cannot be used to detect spread spectrum signals.

B. Matched Filter Detection:

A matched filter is a linear filter which increases the output SNR for an inclined input signal [6]. When primary user's a priori knowledge is known, the Matched filter detection is applied. In matched filter operation the unknown signal is convolved with the filter's impulse response which is time shifted and mirror version of the actual primary user signal. Mathematically, matched filter operation can be expressed as:

$$Y[n] = \Sigma\, h[n-k]\, x[k]$$

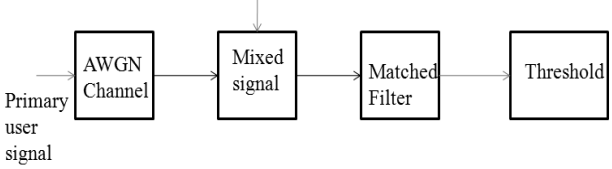


Figure 5: Block diagram of matched filter detection

From the above block diagram, the primary user signal is convolved with time shifted version of the prior noon signal. The final output of the filter is compared with the threshold will determine the primary user presence [7].

C. Cooperative Sensing:

It is used for detecting primary user effectively in high noisy environment, where multiple secondary users make a global decision in relation to primary user[8]. this method is a solution to problems that arise on spectrum sensing due to noise uncertainty, fading and shadowing.

Cooperative sensing techniques:

- Centralized Coordinated
- Decentralized Coordinated
- Decentralized Uncoordinated

Necessisity of cooperative sensing in cognitive radio networks

- Plummeting sensitivity requirements: channel deteriorations like multipath fading, shadowing and building penetration losses, impose high sensitivity requirements inherently limited by power and cost requirements.
- Employing aid between nodes can immensely reduce the sensitivity requirements up to -25 dBm.
- Reduction in sensitivity threshold can be obtained by using cooperation technique.

From above spectrum sensing methods we find some disadvantages like high sensing time taken to achieve a given probability and difficult to differentiate primary user and secondary use signals in energy detection sensing technique, matched filter detection requires prior knowledge of every primary signal and separate receiver for every type of primary user is needed, traffic overhead in cooperative sensing technique. To solve all these problems adaptive spectrum sensing is used.

D. Adaptive Spectrum Sensing:

It is a combination of both single energy sensing and cooperative energy sensing. In this technique SNR is compared with threshold value. If SNR is more single energy sensing is used and if threshold is more cooperative energy sensing is used[6][9].

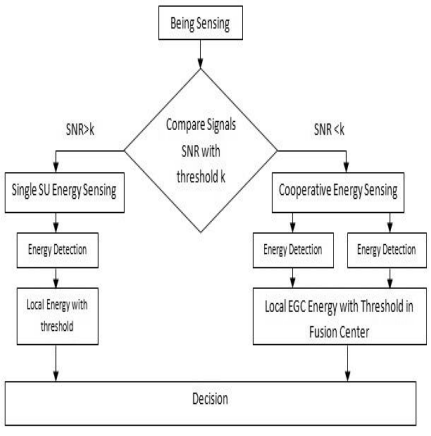


Figure 6: Representation of adaptive spectrum Sensing

V. CONCLUSIONS

The spectrum is a most valuable resource in wireless communication systems and it has been a major research area from last several decades. Cognitive radio is a promising technology which enables spaces. Considering the challenges raised by cognitive radio systems, the use of spectrum sensing methods appears as an essential need to achieve accurate results in terms of efficient use of available spectrum and limited interference with the licensed primary users. In this method, the signal detection is performed by comparing the output of energy detector with a given threshold value. Adaptive spectrum sensing method according to the Secondary users SNR status by selecting both single sensing and cooperative sensing and confirm performance improvement. We also said about importance of cooperative and adaptive cooperative sensing.

REFERENCES

[1] Danijela Cabric, Artem Tkachenko, Robert W. Brodersen,"Spectrum Sensing Measurements of Pilot, Energy, and Col-laborative Detection", Military Communication Conference

[2] Weifang Wang,"Spectrum Sensing for Cognitive radio'Third International Symposium on Intelligent Information Technology Application Workshops, pp: 410-412, 2009..

[3] A.Rahim Biswas, Tuncer Can Aysal, SithamparanathanKandeepan,DzmitryKliazovich, RadoslawPiesiewicz, "Cooperative Shared Spectrum Sensing for Dynamic Cognitive Radio Networks",IEEEInternatiional Conference on Communications (ICC), pp. 1-5,June 2009.

[4] Yi Zheng, XianzhongXie, Lili Yang, "Cooperative SpectrumSensing Based on Blind Source Separation for Cognitive Radio",International Conference on Future Information Network(ICFIN),pp. 398-402, Oct 2009.

[5] Hang Su, Xi Zhang, "Energy-Efficient Spectrum Sensing for Cognitive Radio Networks", IEEE ICC, 2010.

[6] T. Yucek, H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications", IEEE Communications Surveys Tutorials, vol. 11, no. 1, pp. 116–130, 2009

[7] W. Yue and B. Zheng, Spectrum sensing algorithms for primary detection based on reliability in cognitive radio systems, *Journal of Compute and Electrical Engineering*, vol.36, no.3, pp.469-479,2010.

[8] J. Shen, S. Liu, L. Zeng, G. Xie, J. Gao, and Y. Liu, "Optimization of cooperative spectrum sensing in cognitive radio network," Communications, IET, vol. 3, pp. 1170-1178, 2009.

[9] Z. Wei, R. K. Mallik, and K. Ben Letaief, "Cooperative Spectrum Sensing Optimization in Cognitive Radio Networks," in Communications, 2008. ICC '08. IEEE International Conference on, 2008, pp. 3411-3415.

A Simple Power Balancing Algorithm to Improve the Source Utilization Factor for a Multilevel Inverter for Microgrid and Hybrid Electric Vehicle Applications

Phani Kumar K.S.V[1], G.Manohar[2], S.Venkateshwarlu[3]
[1,2,3] Dept of EEE, CVR college of Engineering, Hyderabad, India
Email: phani5016@gmail.com,manohar.gangikunta@gmail.com,svip123@gmail.com

Abstract **—With the promising growth of micro grids and HEVs, the final power quality delivered by an inverter and longevity of the input DC voltage sources has become a subject matter of concern. The conventional multilevel inverter topologies give AC voltages with higher Total Harmonic Distortion (THD) leading to a lower power quality. Also, the unequal usage of the input DC sources leads to a poor Power Source Utilization Factor (PSUF). This paper presents the 3 phase H-Bridge inverter topology with improved line to line power quality features. The proposition of improving Power Source Utilization Factor (PSUF) is achieved by incorporating a novel transposition technique. A comparative study is made to decide the wave shape and minimize the harmonics. The theoretical simulation results and the hardware results for the gate pulses are shown.**

Index terms— **H Bridge Inverter, THD, PSUF, battery balancing algorithm**

I. INTRODUCTION

Various inverters aim to achieve a closer approximation to AC sinusoidal wave by increasing the number of voltage levels. These levels are formed by the utilization of the DC voltage sources and selective sequencing of switches in cascaded mode which tap the source voltage in different proportions to give desired voltage levels at desired time.

There are a number of configurations of multilevel inverters [1] - [4] like voltage-source inverter topologies, including diode-clamped, flying capacitor, and cascaded H-bridge structures. Given the wide range of multilevel inverter topologies, the H-bridged topologies are superior to others as they realize those increased voltage levels for AC approximation levels with lesser hardware[5]-[6].

The paper deals with the inverter architecture, cascading and integrating it to 3 phase, power balancing algorithm and comparison of THDs and PSUFs for different cases has been discussed.

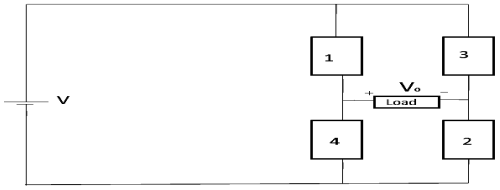


Figure 1. Module of H-Bridge Inverter

II. H - BRIDGE INVERTER

A. H-Bridge Inverters- Architecture

A H-bridge inverter consists of a DC voltage source, 4 switches arranged in a bridge configuration with load connected in between the two arms as shown in the Fig. 1. This is referred to as a module.

B. H Bridge Inverter Operation:

A H-Bridge inverter module powered by a DC voltage source of magnitude V has 3 attainable voltage levels (i.e. +V, 0, -V). These levels are realized using the switching sequence by triggering the gate terminals of the switches desired to be in conduction state. Further it can be observed from Table I. that, at any instant of time, only two switches are ON to give the required voltage level.

The complete voltage appears across the load with V_0 = +V when switches S1 and S2 are turned on. When the switches S3 and S4 are turned on, a total voltage V_0 = –V appears across the load because of the reversal in the current direction. No current passes through the load when all the switches are turned off and V_0= 0.

TABLE-I.

S1	S2	S3	S4	V_0
1	1	0	0	+V
0	0	1	1	-V
0	0	0	0	0

SWITCHING COMBINATIONS OF A SINGLE MODULE OF H BRIDGE INVERTER

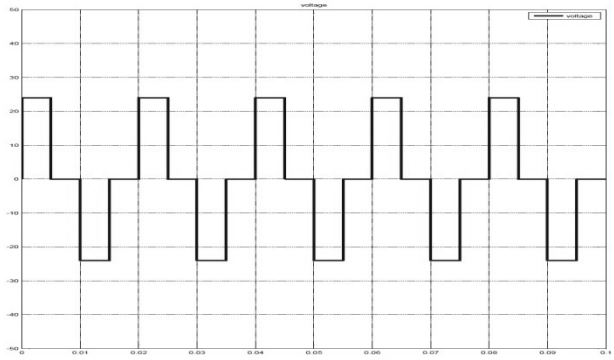


Figure 2. The output wave form of MATLAB simulated model of single module H Bridge inverter showing 3 different voltage levels (+V,0,-V).

III. CASCADING OF MODULES

A. Single Phase Cascaded H Bridge Inverter

Cascading is the series addition of a new module to an existing module. It increases the peak to peak voltage thereby creating new accessible voltage levels. The step increase in voltage level is equal to the magnitude of the voltage source in a single module of H-Bridge Inverter. As cited in [7] the number of accessible voltage levels is a function of the number of modules being cascaded

$$L = 2n + 1 \quad (1)$$

Where L = Number of different voltage levels
n = number of modules

Thus for three modules of cascading (as shown in Fig.3) each with a DC source of 'V' volts, the number of different voltage levels range from 3V to -3V covering a total of seven levels as shown in Fig. 4.

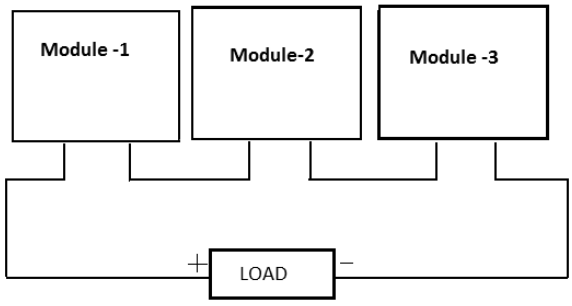


Figure 3. Single phase H Bridge inverter cascaded thrice

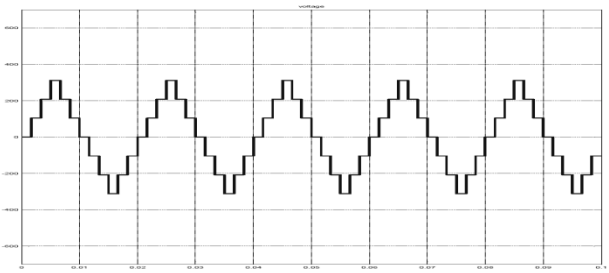


Figure 4. The output wave form of MATLAB simulated model of 3 module cascaded H bridge inverter showing 7 different voltage levels.

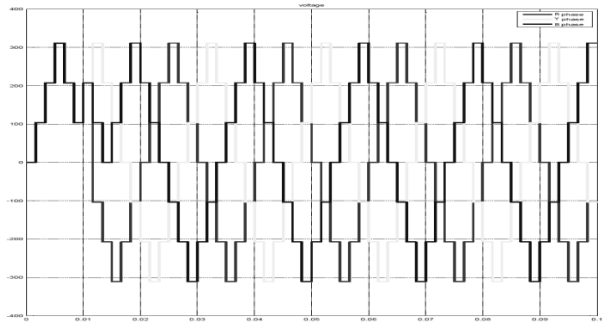


Figure 5. The output wave form of MATLAB simulated model of 3 phase (RYB) cascaded H bridge inverter showing 7 different voltage levels.

B. Three Phase Cascaded H-Bridge Inverter

The cascaded modules are integrated to three phase cascaded HBridge inverter and given to a three phase Delta connected load as it eliminates third order harmonics. The three phases are independant of each other and displaced by 120^0 as shown in the Fig 5.

IV. POWER BALANCING ALGORITHM

In case of conventional H Bridge inverters the power sources in 'n' level cascaded unit are not utilised equally. This leads to an unequal Power Source Utility Factor. To equalize the PSUF a specific power balancing algorithm has been proposed. For an 'n' cascaded module the number of cycles considered are 'n'. In these n cycles, the power being fed by the power source of each module is equalized using transposition technique which employs selective switching patterns fed to the gate terminals of switches as pulses.

Here comparative study is made between cases of wave-shape. Case – 1 is a wave-shape with two peak voltage levels and two zero voltage levels per cycle. Case – 2 corresponds to two peaks and zero zeros per cycle. Case – 3 deals with four peaks and zero zeros per cycle.

For example, consider a 3 phase 3 module multilevel H bridge inverter where the analysis is made for PSUF with and without transposition the PSU values for each cycle are analyzed as follows:

Consider three cycles of the output voltage waveforms as shown in Fig. 6 and Fig. 7 where the alphabet represent the module, positive sign represents the switching on of a switch and negative sign represents that the switch is off or acts as a bypass.

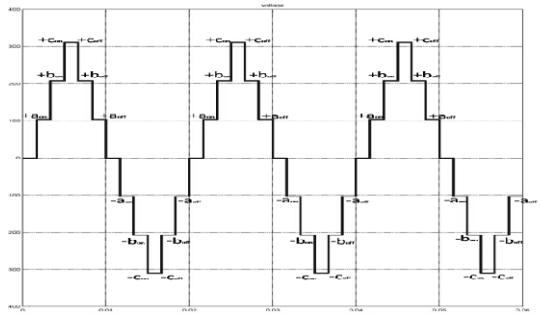


Figure 6. Output waveform of cascaded H-Bridge inverter of Case-1 without transposition.

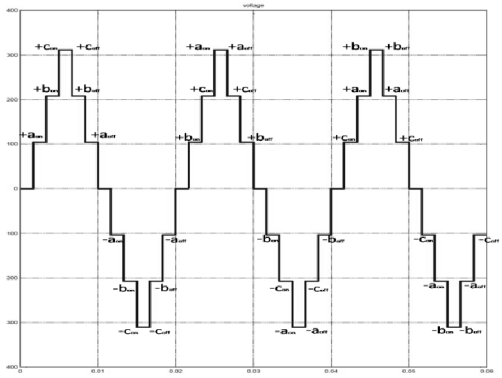


Figure 7. Output waveform of cascaded H-Bridge inverter of case-1 with transposition

TABLE II

COMPARISON OF PSUF FOR CASE - 1

		Without transposition			With transposition		
		MODULES			MODULES		
		1	2	3	1	2	3
CYCLES	C1	10/12	6/12	2/12	10/12	6/12	2/12
	C2	10/12	6/12	2/12	2/12	10/12	6/12
	C3	10/12	6/12	2/12	6/12	2/12	10/12
	PSUF	30/36 = 83.33 %	18/36 =50 %	6/36 = 16.67%	18/36 = 50%	18/36 = 50%	18/36 = 50%

The sequence for without transposition and with transposition is shown in the Fig. 6 and Fig. 7 respectively. The analysis is shown in the Table II. PSUF when not transposed is highest for the voltage source in Module 1 and lowest for the voltage source in Module 3. Whereas, when the switching pulses are transposed, the values of PSUF for all the modules are equal and hence power utility is balanced.

Similarly, other output wave forms and corresponding tables comparing the transposed and un-transposed PUSF for 2 peaks 0 zeros are Fig. 8 and Table III respectively, and for 4 peaks 0 zeros are Fig. 9 and Table IV respectively.

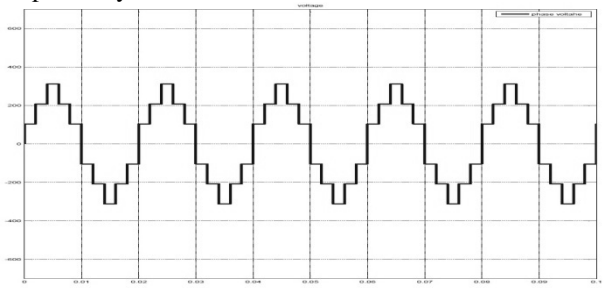
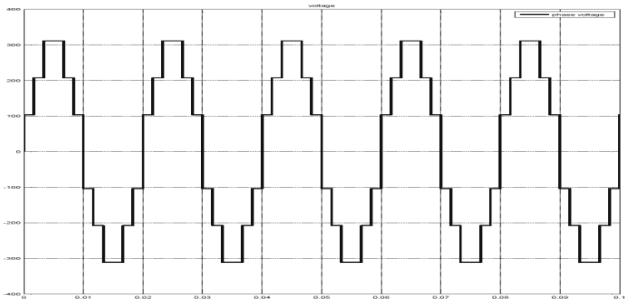


Figure 8(a) The wave form of Case -2

TABLE III

COMPARISON OF PSUF FOR CASE-2

		Without transposition			With transposition		
		MODULES			MODULES		
		1	2	3	1	2	3
CYCLES	C1	10/10	4/10	2/10	10/10	4/10	2/10
	C2	10/10	4/10	2/10	2/10	10/10	4/10
	C3	10/10	4/10	2/10	4/10	2/10	10/10
	PSUF	30/30 = 100%	12/30 = 40%	6/30 = 20%	16/30 = 53.33 %	16/30 = 53.33 %	16/30 = 53.33 %



Figures 8(b) The wave form of CASE - 3

TABLE III

COMPARISON OF PSUF FOR CASE - 3

		Without transposition			With transposition		
		MODULES			MODULES		
		1	2	3	1	2	3
CYCLES	C1	12/12	8/12	4/12	12/12	8/12	4/12
	C2	12/12	8/12	4/12	4/12	12/12	8/12
	C3	12/12	8/12	4/12	8/12	4/12	12/12
	PSUF	36/36 = 100%	18/36 =66.67 %	12/36 = 33.33%	24/36 = 66.67 %	24/36 = 66.67%	24/36 = 66.67 %

V. SIMULATION RESULTS AND HARDWARE

A.Simulation

As per the section 3.B a simulink model has been developed as shown in the Fig. 9

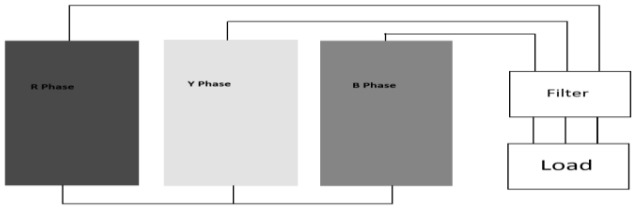


Figure 9 Block diagram of thre phase cascaded three module H Bridge inverter.

The line to line output voltage waveform is as shown in the Fig.10

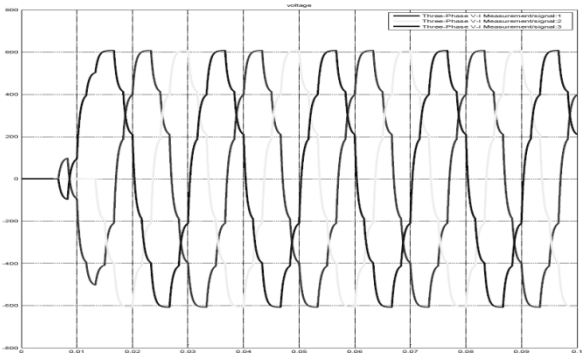


Figure 10. Line to line output voltage of three of thre phase cascaded three module H Bridge inverter.

The pulse for first switch of module 1 in phase A are developed for two cases i.e., with transposition and without transposition as shown in Fig.11 and 12 respectively.

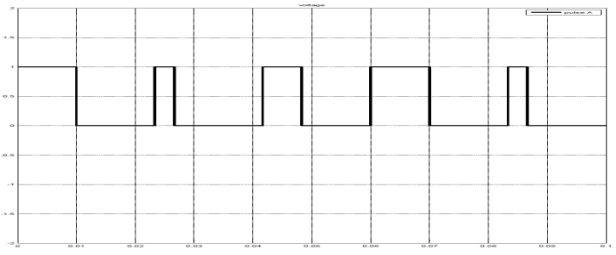


Figure 11. Gate pulse for transposed

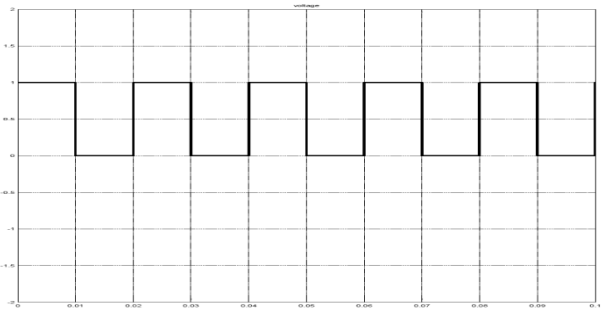


Figure 12. Gate pulse for un-transposed conventional

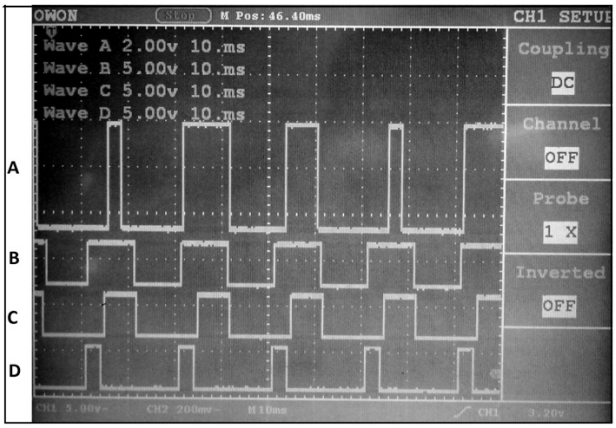


Figure 13. Transposed gate pulse (A) versus conventional gate pulses (B, C, D) generated and plotted on a CRO

The switch pattern generated by simulink has been interfaced to ATMEGA _ (Arduino UNO) so that the micro controller acts independently thereafter the pin output of the microcontroller produces the waveform as shown in the Fig. 13.

In waveform 'A' from b^{th} – $(b+0.02)^{th}$ (s) the source shall be ON for 16.67% of 0.02 seconds as shown in the fig. 13.D. In $(b+0.02)^{th}$ - $(b+0.04)^{th}$ second it conducts for 50% of time as shown in the Fig. 13. C. and in $(b+0.02)^{th}$ - $(b+0.04)^{th}$ second, it conducts for 83.3% of time as shown in the Fig. 13. B. where B, C, D are the pulses to switch 1 in modules 1, 2, 3 without transposition and 'b' is the integral multiple of 0.06 (i.e. b = 0.06*i) where 'i' is an integer.

VI. COMPARISION OF THD

The THD content in a wave is calculated after each waveform is simulated in MATLAB 2013. The theoretical calculations involved in THD computation are elucidated in [8]. A comparison of THDs of different cases gives the following results as shown in the Table IV

TABLE IV
COMPARISON OF THDs

CASE	THD
CASE – 1	5.09%
CASE – 2	7.1%
CASE - 3	4.97%

The figure to show the minimum THD obtained in Case-3 is shown in the Fig. 14

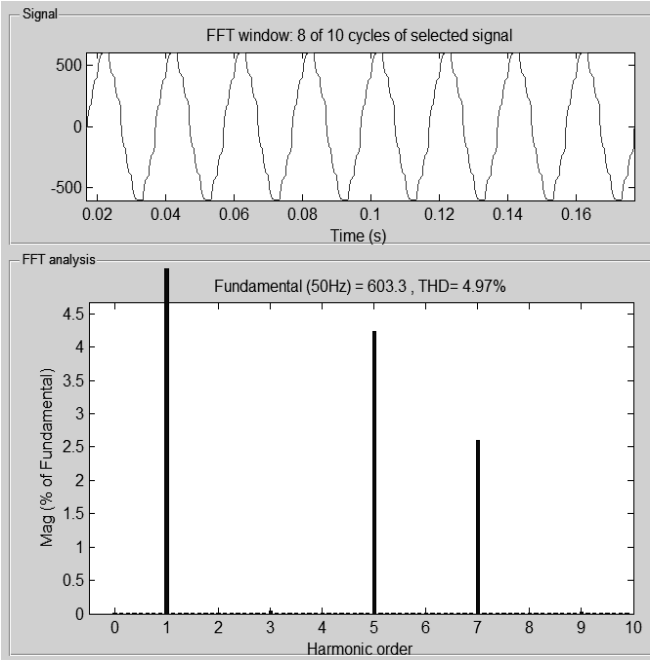


Figure. 14 THD of 4 peak 0 zero

VII CONCLUSIONS

The Total Harmonic Distortion of line-to-line voltage in a three phase H Bridge for different cases (1, 2 and 3) has been compared and found that cases 1 and 3 comply to the IEEE standards. The Power Source Utility factor in different modules of a cascaded unit has been improved significantly. This power source balancing has been realized by practically generating the pulses.

REFERENCES

[1] S. B. Kjaer, J. K. Pedersen, and F. Blaabjerg, "A review of single-phase grid-connected inverters for photovoltaic modules," *IEEE Trans. Ind. Appl.*, vol. 41, no. 5, pp. 1292–1306, 2005.

[2] S. Jain and K. . Gupta, "Topology for multilevel inverters to attain maximum number of levels from given DC sources," *IET Power Electronics*, vol. 5, no. May 2011, pp. 435–446, 2011.

[3] N. Yousefpoor, S. H. Fathi, N. Farokhnia, and H. A. Abyaneh, "THD minimization applied directly on the line-to-line voltage of multilevel inverters," *IEEE Trans. Ind. Electron.*, vol. 59, no. 1, pp. 373–380, 2012.

[4] S. Reddy, "A Review And Harmonic Analysis Of Modified H-Bridge Multilevel Inverter Fed Indution Motor Drive," *Int. J. Eng.*, vol. 1, no. 7, pp. 1–12, 2012.

[5] J. Rodriguez, J. S. Lai, and F. Z. Zeng, "Multilevel inverters: A survey of topologies, controls and applications," IEEE Trans. Ind. Electron., vol. 49, no. 4, pp. 724–738, Aug. 2002.

[6] R.Shelar S.M.Chaudhari "Speed Control of 3 phase of Induction Motor by Using Multilevel Inverter" International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163 Volume 1 Issue 12(December2014

[7] Rajesh Kr Ahuja, Lalit Aggarwal, Pankaj Kumar, "Simulation of Single Phase Multilevel Inverters with Simple Control Strategy Using MATLAB" International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 2, Issue 10, October 2013.

[8] Gobinath.K, Mahendran.S, Gnanambal.I, "New cascaded h-bridge multilevel inverter with improved efficiency" International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering. *Vol. 2, Issue 4, April 2013*

A Solar Photo-Voltaic and Battery Integrated Microgrid Test System Performance Analysis

Phani Kumar K.S.V[1], S.Venkateshwarlu[2]
[1,2]Dept of EEE, CVR college of Engineering, Hyderabad, India
Email: phani5016@gmail.com, svip123@gmail.com

Abstract— **Power crisis has been a matter of major concern with losses in the transmission adding to the woes. The former can be overcome by replacing the conventional energy sources (CES) with the renewable energy sources (RES). And the latter problem can be dealt by using Microgrid technology which is taking boom. This paper deals with the integration of renewable energy sources with the Microgrid technology considering the standard IEEE Fourteen bus system as reference. The entire system is simulated, operation is analyzed and the final results have been published.**

Index terms—**Microgrid, IEEE Fourteen bus system, RES, PV array**

I. INTRODUCTION

The energy sources used for the present power generation are getting exhausted day by day. So, the usage of renewable energy sources would be of great importance. Also, the available transmission and distribution efficiency of 16.45% in AP is not up to the mark.

A small energy system capable of generating power within a defined boundary integrated with variable Distributed Energy Resources (DERs) called Microgrid (as stated in IEEE Std 1547.4-2011 [1]) seems to be a perfect solution for overcoming these problems. Some of the resources that can be integrated with microgrid are solar energy (using the photo voltaic systems), wind energy and bio-mass energy, etc. Depending on the availability of the source and the cost of production of power the choice of selection of a distributed source is made for the power generation. There are two catalysts boosting the renewable energy integration into the grid. Firstly, the cost of power production is decreasing as the investment cost per KW of production is decreasing as bulk manufacturing is happening and many players have come into the market competing for the ever increasing customer base.

Much of the research is being done in operating a microgrid in two modes i.e., grid connected mode, islanding mode [2]. Based on the measurement and information systems, there can be two modes of control, i.e., Centralized and distributed modes [3] complying with IEC61850 standards. It provides service reliability and reduces the cost of energy encouraging the usage of eco-friendly renewable energy resources. The simulation of microgrids has been traditionally done in system level and component level. The system level simulation yields results that can be used in the stability, reliability and power quality analysis. The simulation results are further compared to the grid standards and the choice of controlling action is chosen accordingly. The concept of net metering has brought many security issues into light and the Indian Electricity act 2003 highlights the rules for grid connectivity and the sale of electricity in section 86 (1) (e) [4]. In India, the policies, initiatives, subsidies, research and development are significant for the growth of solar photovoltaic power generation for roof-top off-grid, net-metered rooftop and concentrated solar power. The solar map of India also proves the extensive availability of solar energy [5].

From 1859, many chemical combinations of rechargeable lead-acid batteries were developed and ever since then they have been used for energy storage. Though Sony has made lithium-ion batteries commercially available from 1991, the market has not responded so well in India. Some pilot projects like Sendai Substation Lithium Ion Battery Pilot Project [6] is under construction in Japan for 30 min duration at rated power, with a capacity of 20 MWh and maximum output of 40 MW.

The presence of a storage system enhances the reliability of the system but poses many security threats and controlling the system. Battery Management is also another area that has been researched about and shall be included in the later work as an extension to this paper work.

The rest of the paper is organized as following sections:

1) Objective: This section briefs about the objective of the paper.
2) System structure: This section presents the structure of the system which has been considered in this paper and the details pertaining to it.
3) Solar Energy: This part provides the brief idea of the working of solar cells.
4) PV modelling: This part presents the equivalent circuit of a PV cell and the related equations.
5) Inverters: This topic deals with the basic types of inverters and the inverter which is being considered in this paper (Cascaded H-bridge type).
6) Solar grid interfacing: This topic deals with the challenges faced in the interconnection of Distributed Electric Production resources with the utility grid and the international standard governing them.
7) Simulation: The results of simulation have been presented in this section.
8) The interferences made from the results have been summarized in this section.

II. OBJECTIVE

This paper deals with testing of the feasibility of implementing the DERs in a microgrid system and to study the performance of the system, considering the standard IEEE fourteen bus system as reference. Solar Energy is taken as the source of renewable energy and the conventional synchronous generators are replaced with solar arrays coupled with a cascaded H-bridge inverter. The main objective is to obtain the voltage at load bus and verify it with the existing standard.

III. SYSTEM STRUCTURE

The test system (Fig.1) which has been considered consists of:

a) Fourteen buses
b) Two solar power generators in buses 1 and 2.
c) Three synchronous compensators to provide reactive power compensation in buses 3, 6 and 8.
d) loads in buses 2, 3, 4, 5, 6, 9, 10, 11, 12, 13 and 14

The line and bus data have been provided in the appendix.

IV. SOLAR ENERGY

Solar energy is one of the abundantly available renewable sources on earth. Due to its no-polluting nature, it is being used in major applications involving energy generation. The best way to harness this solar energy would be through the photovoltaic cells. They convert solar energy into DC electricity, which can be used for various purposes. A photovoltaic (PV) cell consists of a P-N unction. Whenever light falls on a PV cell, if the incident energy is greater than or equal to the bonding energy of the electron, then an electron hole pair is generated. The electron which is majority charge carrier (in case of N- layer) passes through the connected load and meets the hole which has crossed the junction. In this way, the circuit is completed and electricity is generated. The general formula used to estimate the output electricity of photo voltaic system is given by

$$E=A*t*r*H*PR \quad (1)$$

Where, E= energy (KWh)
A=total area of the solar panel,

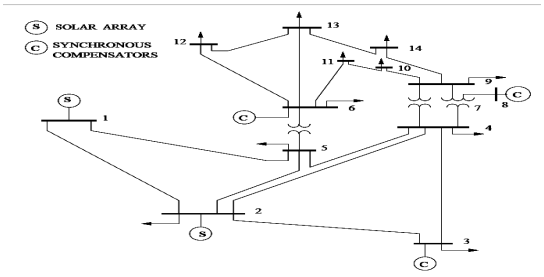


Figure 1. IEEE 14-Bus System With Solar Power Generators

r = yield of the solar panel,
H=annual average solar radiation on fitted panels,
PR=performance ratio (coefficient of losses)

V. MODELLING OF SOURCES

A. PV Module

The solar cell can be considered as a type of p-n junction diode. Unlike a conventional diode, the electron-hole (e-h) pair in a solar cell is formed by photovoltaic effect. The e-h pairs are separated by the electric field existing across the junction and are driven through the external load by the junction potential.

The solar cell can be considered as a current source. The current supplied by the solar cell gets divided as the load current and the leakage current. Thus, the losses can be accounted by considering two resistances: series and shunt. The reverse saturation current also adds to these losses. Considering the above factors into account, the equivalent circuit of the solar cell can be drawn as shown in Fig. 2 below.

The output terminal current can be given by the equation (2)

$$I = I_{ph} - I_D - I_{sh} \quad (2)$$

Where, I is the load current
I_{ph} is the total light generated current
I_D is the diode current
I_{sh} is the shunt leakage current
In an ideal solar cell, $R_s=0$ and $R_{sh}=\infty$.

Small changes in shunt resistance will not have much effect on the efficiency but, small changes in the series resistance can cause the PV output to reduce significantly.

The load current of the solar cell is given by the expression (3)

$$I = I_{ph}-I_{os}\{\exp[qU_{oc}/AKT]-1\}-(U_{oc}/R_{sh}) \quad (3)$$

Where,
I_{ph} = total light generated current (A)
I = cell output current (A)
I_{os} = cell reverse saturation current. (A)
q = electron charge=1.6*10^19 Coulombs.
A = ideality factor of p-n junction.
K = Bolzmann constant.
T = cell temperature [^oC].
R_{sh} = shunt resistance. (Ohms)
U_{oc} = terminal voltage of the cell (V)

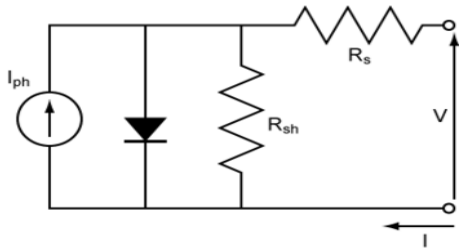


Figure 2. Equivalent Circuit of PV cell

B. Battery Module

A typical model of lead acid battery can be understood in two equations,(4,5) corresponding to the charging and discharging models. The flow of current while charging is

considered to be negative for charging and positive for discharging. The models can be written, with reference to the MATLAB as:

Discharge model (i* > 0)

$$f_1(it,i^*,i,Exp) = E_0 - K\cdot\frac{Q}{Q-it}\cdot i^* - K\cdot\frac{Q}{Q-it}\cdot it + \text{Laplace}^{-1}\left(\frac{Exp(s)}{Sel(s)}\cdot 0\right). \quad (4)$$

Charge Model (i* < 0)

$$f_2(it,i^*,i,Exp) = E_0 - K\cdot\frac{Q}{it+0.1\cdot Q}\cdot i^* - K\cdot\frac{Q}{Q-it}\cdot it + \text{Laplace}^{-1}\left(\frac{Exp(s)}{Sel(s)}\cdot\frac{1}{s}\right). \quad (5)$$

Where,

E_{Batt} = Nonlinear voltage (V)
E_0 = Constant voltage (V)
Exp(s) = Exponential zone dynamics (V)
Sel(s) = Represents the battery mode. Sel(s) = 0 during battery discharge, Sel(s) = 1 during battery charging.
K = Polarization constant (Ah^{-1})
i* = Low frequency current dynamics (A)
i = Battery current (A)
it = Extracted capacity (Ah)
Q = Maximum battery capacity (Ah)
A = Exponential voltage (V)
B = Exponential capacity $(Ah)^{-1}$

VI. INVERTER

Inverters are the devices which convert DC voltage to AC voltage. There are two basic types of inverters: grid interactive inverters and standalone inverters. Grid interactive type inverters are supplied with voltage and frequency by the grid itself and they cannot operate without the supply from the grid. Standalone inverters can operate without a grid and can serve as a back-up for the renewable generation.

The topology used here is a three-phase cascaded multi-level inverter (without battery balance). It basically consists of series connected modules consisting of a H-bridge. Single phase version of the topology is shown in the Fig. 3.

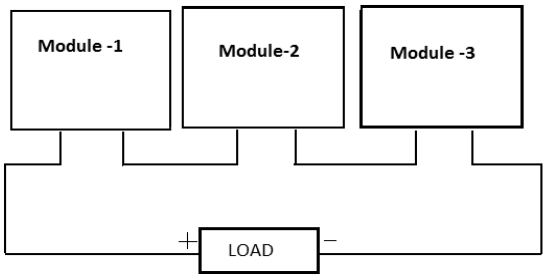


Figure 3. Block Diagram Of Cascaded H-Bridge Inverter

To get a sinusoidal wave, the switching sequence is appropriately controlled. With a single H-bridge inverter, we get three levels (two non-zero levels and a zero level) in the output waveform. To make the waveform more sinusoidal in nature, we cascade more number of H-bridge inverters, which increases the non-zero levels obtained in the output waveform. In this paper, solar array is considered as the isolated dc source that the topology demands. The switching frequency used here is the fundamental frequency so that the switching losses are reduced while the Total Harmonic Distortion (THD) is observed not to be deviating from the standards to be maintained at the Point of Common Coupling (PCC).

VII. SOLAR GRID INTERFACING

The power quality of a solar panel is affected by various parameters including irradiation, PV modules, inverters, etc. A small change in the irradiance level and cloud cover or shading effects play a vital role for low-voltage distribution grids with high penetration of photo-voltage, as presented in [2].

These voltage fluctuations can cause the disconnection of inverters from the grid. Also, in the long run, the performance efficiency of the grid-connected PV systems reduces due to the source variation and inverter performance [3].

Therefore, attention has to be given to the voltage and power profiles.

The general block diagram of a single-phase or three-phase PV system is as shown in the Fig. 4. The PV array can be a single panel or a series and parallel combination of various PV modules. The PV systems can also be used in centralized and decentralized modes and the summary of advantages and disadvantages of these topologies are discussed in [4].

The major factors to be considered while designing the inverter are the injection of harmonics due to the usage of inverter, irradiance and the shading effects [5-6]. In this paper, the cascaded H-bridge inverter is being used.

There are a number of technical challenges associated with the parallel connection of distributed electric production resources with the utility grid which include frequency and voltage regulation, disconnection and reconnection rules during the disturbances in the grid, islanding operation and faults minimization when connected to grid.

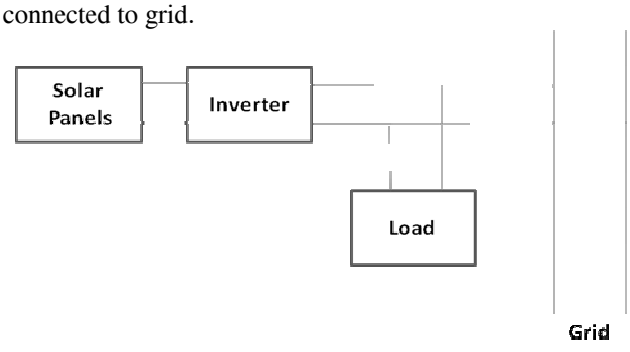


Figure 4. General Structure Of Grid-Connected PV System

A. International Standards:

According to IEEE standards, Electric Power Stations (EPS) refer to the Utility Grids while Distributed

Resources (DRs) include mini-grid systems. Utility interconnection of electric distributed resources is specified by standards as mentioned in table 1 as:

TABLE 1. STANDARDS FOR MICROGRID

IEEE 1547	The guidance related to the operating island systems in various modes is provided [7]
IEC	Microgrid and their safety concerns are addressed.

In India, 'smart grid forum's working group #9 on renewable and microgrids' is developing standards for the integration of renewable-based microgrids with the main grid.

India's Central Electric Authority released a set of standards for distributed generation resources connectivity [8] in 2012.

VIII. RESULTS

The generator of the fourteen bus system has been replaced by a unit consisting of three solar arrays per module per phase. The properties of them can be listed as

Module type : Canadian solar CS5P-220M
No. of cells per module : 96
No of series connected modules per string: 1
No of parallel strings : 1
Module specifications under STC
Voc : 59.2 V
Isc : 5.09 A
Vmp : 48.3 V
Imp : 4.54 A
Sample time : 1.00E-06

The characteristics of the solar module used can be shown as:

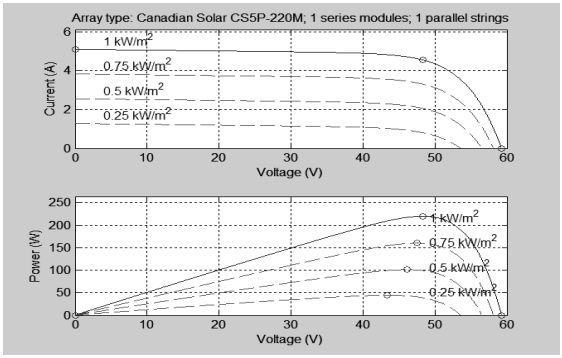


Figure 5. I-V, P-V Characteristics Of Array at 250C.

The solar irradiation as a function of time taken for the simulation is plotted and displayed in fig 6. as

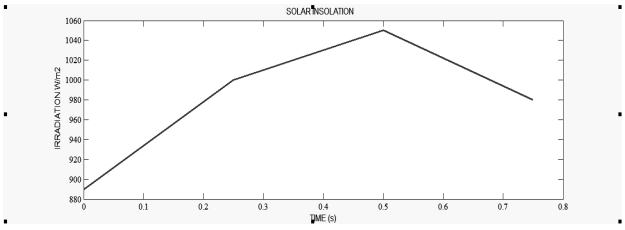


Figure 6. Insolation Plot

Similarly, the battery used as here a lead-acid battery with the following parameters:
Nominal voltage : 50 V
Rated Capacity : 6.5 A
Initial State of charge : 100 %
Maximum capacity : 8 Ah
Fully charged Voltage : 52V
Nominal Discharge Current: 5.5 A
Internal Resistance : 0.0018462
Capacity @ Nominal Voltage: 6.25 Ah

The DC output of the solar module is shown in figure-7 as:

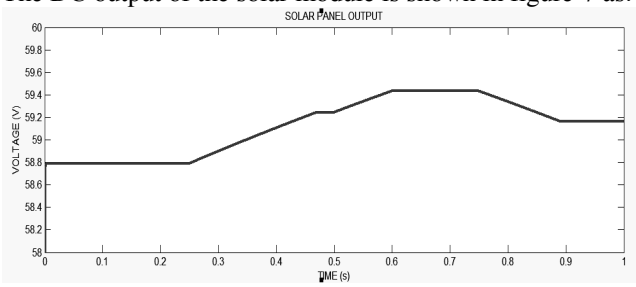


Figure 7. DC Output Of The Solar Module

The Single phase output of the multilevel inverter is shown in figure-8 as:

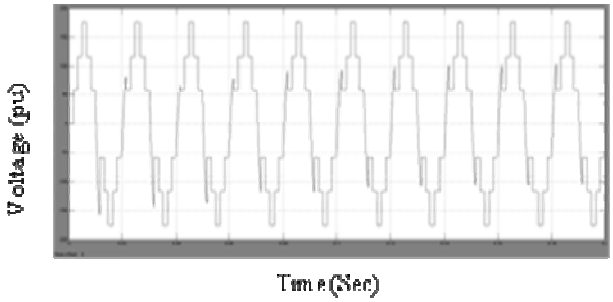


Figure 8. Single Phase Output Of The Multilevel Inverter

The Phase voltage of the three phase PV system is shown in figure 9 as:

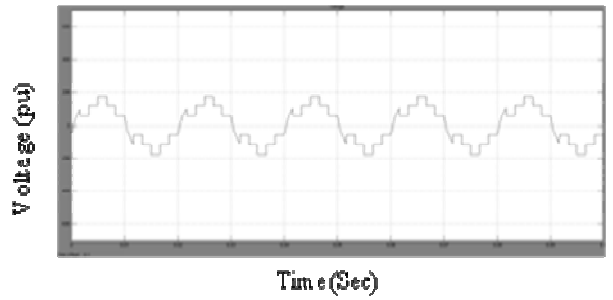


Figure 9. Phase Voltage Of The Three Phase PV System

The Three phase RMS voltage at generator bus (bus 1) and load bus (bus 5) respectively are shown in figure-10 as:

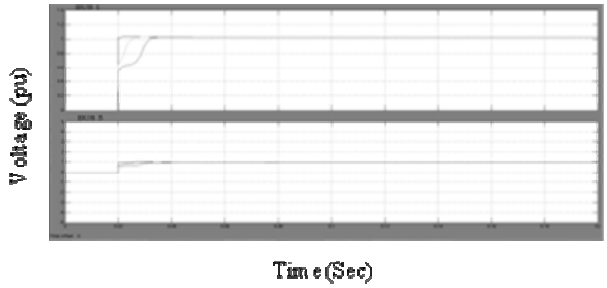


Figure 10. Three Phase Rms Voltage At Generator Bus (Bus 1) And Load Bus (Bus 5)

A. Result:

The bus voltages at all the buses are listed in the table 2:

TABLE 2. BUS VOLTAGES IN PER UNITS

Bus No.	Voltage (pu)	Bus No.	Voltage (pu)
1	1.015143527	8	1.012934649
2	1.040391553	9	0.98398143
3	1.029143689	10	0.984292008
4	1.018104473	11	1.004629565
5	1.022478674	12	1.021498803
6	1.035501198	13	0.97822907
7	1.00063217	14	0.979625316

IX. CONCLUSIONS

The paper has presented the integration of Renewable energy source with the microgrid system, considering the standard IEEE fourteen bus system data as reference. The result has been analyzed and summarized and it has been shown that the voltage values obtained is within the ±5% tolerance of the standard voltage value (1 pu).

REFERENCES

[1] C. M. Colson, S. Member, M. H. Nehrir, and S. Member, "A Review of Challenges to Real-Time Power Management of Microgrids," *IEEE Power Energy Soc. Gen. Meet.*, no. 978–1–4244–4241–6/09, pp. 1–8, 2009.

[2] IEEE Standards Coordinating Committee 21, "IEEE Std.1547.4-2011: IEEE Guide for Design, Operation, and Integration of Distributed Resource Island Systems with Electric Power Systems", Jul. 2011

[3] C. M. Colson, S. Member, M. H. Nehrir, and S. Member, "A Review of Challenges to Real-Time Power Management of Microgrids," *IEEE Power Energy Soc. Gen. Meet.*, no. 978–1–4244–4241–6/09, pp. 1–8, 2009.

[4] N. K. Sharma, P. K. Tiwari, and Y. R. Sood, "Environmental Friendly Solar Energy in Restructured Indian Power Sector," *Int. Conf. Sustain. Energy Intell. Syst.*, vol. 2, no. June 2008, pp. 104–109, 2012.

[5] http://en.wikipedia.org/wiki/Solar_power_in_India

[6] http://www.energystorageexchange.org/projects/ global_search?q=sendai.

[7] Institute of Electrical and Electronics Engineers, "IEEE Guide for Design, Operation, and Integration of Distributed Resource Island Systems with Electric Power Systems," IEEE Std 1547.4, Jul. 20, 2011.

[8] Central Electricity Authority (India), "Technical standards for connectivity of the distributed generation resources." 2012.

Implementation of Open Source IP based Embedded System

Y.Divya reddy
Department of EIE, CVR College of Engineering, Hyderabad, India
Email: divya.reddy037@gmail.com

***Abstract*—Embedded system plays an important role in various industry applications. An embedded system is consisting of software and hardware. The hardware platform of conventional embedded system is typically based on IC chips that have fixed resources. Besides, with the development of FPGA, an emerging approach for designing embedded system is implementing soft IP cores on FPGAs. Soft IP cores are synthesizable hardware blocks described in HDL language. Their source code can be either open or close to public. For example, Open RISC 1200, is an open source 32-bit RISC microprocessor. In addition, the increasing complexity of embedded system forces software developers to consider operating system support to reduce their workload. Thus, in this paper, a prototype of open source IP based embedded system with Linux is implemented on Atlys (Xilinx Spartan-6) FPGA board and the goal is to evaluate if the system is appropriate for industrial applications. The hardware platform is ORPSOC, which is a reference SoC design based on Open RISC 1200 processor. For software, Linux operating system is installed. Furthermore, an application executes on Linux is developed that reads the output of an I2C compass sensor-LSM303DLM. With the success of the application and the investigation of license issues, the conclusion is drawn that open source IP based embedded system with Linux is usable for industry. Although comparing to conventional embedded system, the open source IP based embedded system with Linux has following cons, such as high product cost, basic-supported development environment and more difficult software development if Linux driver doesn't support the hardware. However, its pros are high flexibility and scalability, high software portability, low software development difficulty and high reusability that make it more suitable for industry usage.**

***Index Terms* - embedded system, open source hardware, Open RISC 1200, Linux**

I. INTRODUCTION

The embedded system plays a more and more important role in the world. It can be found in many applications for industrial use, such as process control and monitoring systems. The task for an embedded system is typically to control the machine to function correctly. It can be said that the industrial productivity is improved with introducing embedded system. Differs from general purpose computer, an embedded system is designed for a dedicated task, which requires both hardware and software operate appropriately. Normally, the embedded system designed in conventional way is consisting of 3 layers as shown in Figure 1. The hardware platform is a PCB board containing a microcontroller and other peripheral ICs such as on-board memory, Ethernet module and etc. The software is the application code of the desired task including hardware drivers.

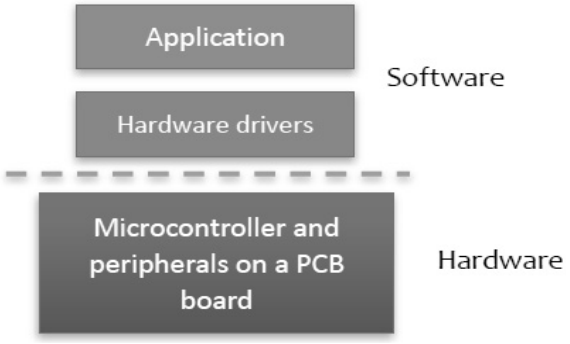


Figure 1: Architecture of conventional embedded system

With the rapid development of semiconductor technology, the capability of embedded system is increasing over time to allow more complex task to be performed. Thus, the embedded operating system support becomes necessary. Figure 2 illustrates the architecture of a conventional embedded system with Linux. The benefits of introducing Linux are providing various hardware drivers, communication protocols and management of system resources that reduce the workload of software developer significantly.

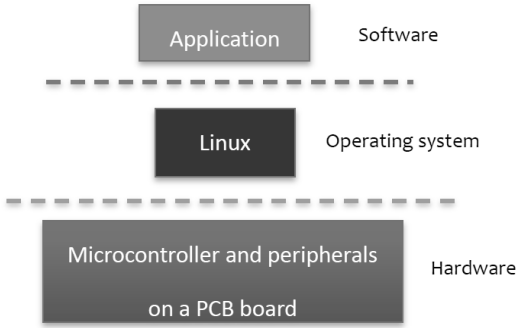


Figure 2: Architecture of conventional embedded system with Linux.

In addition, FPGA (Field Programmable Gate Array) is an alternative for embedded design. It is well known as its hardware re-programmable feature. With this advantage, an emerging technology for achieving embedded system's hardware platform is using soft IP cores. The term IP refers to intellectual property. Soft IP cores are the synthesizable hardware blocks for FPGAs and described in RTL level. A soft IP can be either a processor or other hardware modules such as UART, ETHERNET and I2C controller. Designing embedded system with soft IP cores is flexible and

configurable. At present, there are two leading-edge commercial soft processors, Microblaze and Nios II, supported by Xilinx and ALTERA. Besides, there are other open source processors, such as Open RISC and LEON3. Open IP cores inherit the advantages of soft IPs and are open source to public, which means they are free, and FPGA independent. Therefore, it is meaningful to implement an open source IP based embedded system with Linux to evaluate if it is compatible for industrial applications. In this paper, the hardware platform is ORPSOC (Open RISC Reference Platform System-On-Chip), which is a reference, embedded system design based on Open RISC soft processor. Then the Linux operating system is installed and an application running on Linux that reads the acceleration raw data from an I2C compass sensor, LSM303DLM.

II. Open Source Hardware

The commercial soft IP cores such as Microblaze and Nios II are leading a revolution of embedded system design; they are high performance, well supported and flexible. However, they are not open source and FPGA dependent that means the designers only has the right to use them in their embedded design and the implementation of soft IP cores is only possible on specific vendor's FPGAs. In software world, open source software is becoming popular. The "open source" means "freedom" that everyone has the right to modify it and derived works are allowed under the condition that the developers should pass on the freedom to others. This causes a great success of open source software such as Linux operating system. In hardware world, modern silicon chip is typically built from silicon "intellectual property" (IP), written in a hardware description language. Fabless design houses may never produce a chip themselves—one of the largest and best known is ARM in Cambridge, whose processor IP is built by other companies into one billion chips ever month. That IP costs the same amount to produce, whether it goes into one chip or one billion.

A. Open RISC processor and ORPSOC

Open RISC CPU architecture, one of the flagship projects of opnecores.org is a well-known open source processor. Open RISC 1200 processor is an implementation of Open RISC 1000 processor family. Figure 3 shows the architecture of Open RISC 1200 CPU.

The Open RISC 1200 CPU is a 32-bit scalar RISC with Harvard micro architecture, 5 stage integer pipeline, virtual memory support (MMU) and basic DSP capabilities. The MMU enables the capability of running an operating system. Supplemental facilities include debug unit for real-time debugging, high-resolution tick timer, programmable interrupt controller and power management support[2]. This processor can be synthesized and downloaded onto Altera and Xilinx FPGAs and supports embedded real time operating systems such as Linux, µLinux and OAR RTEMS real time operating system. For software development, tools are available that allow developers to compile programs written in C/C++, Java and Fortran to run on the Open RISC processor [1].

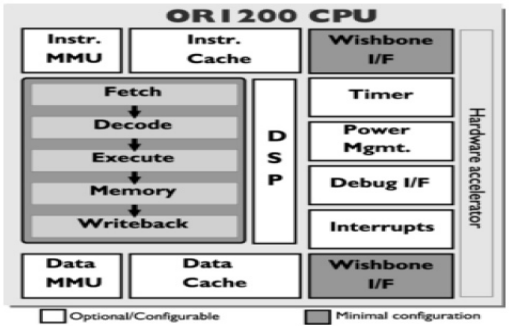


Figure 3: Open RISC 1200 CPU architecture

ORPSOC (Open RISC Reference Platform System-on-Chip) is a reference system-on-chip design that is primarily for Open RISC based embedded system testing and development. And it is also the hardware platform of this thesis that contains an Open RISC 1200 processor and several peripherals. The interconnections between CPU and peripherals are using Wishbone interface, which have been configured already. In ORPSOC's source code, there are make file scripts to generate FPGA configuration bit stream for particular FPGA boards, such as selected Atlys FPGA board in this paper.

B. Linux user space and kernel space

Linux is a successful open source operating system that is widely used in embedded systems as a platform for executing applications. It manages the resources of an embedded system that the software developer can focus on application code on a high-level abstraction view without being involved in hardware driver development if the embedded system performs a complex task. In Linux, memory is divided into two spaces, one is user space, and the other one is kernel space. Figure 4 shows the relationship between user space and kernel space.

The top is user space where applications execute, while the kernel space is an exclusive space only for kernel running. The kernel has the highest authority to access all resources in an embedded system, such as memory and devices. And it should be as stable as possible to prevent any undesired errors happening and coordinates processes. By contrast, user space has less authority that it can't access the data in kernel space directly. The data transition between user space and kernel space is via system call interface.

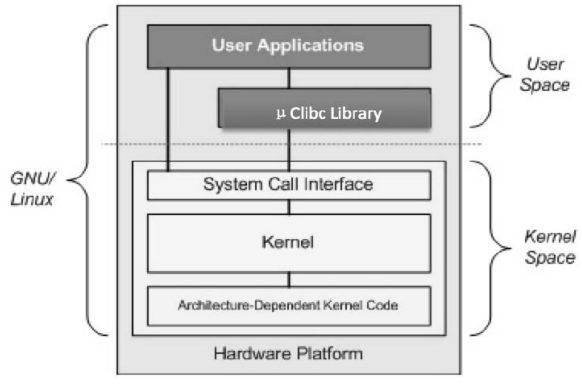


Figure 4: user space and kernel space

Once a user application invokes a system call to access a particular hardware, the kernel handles it with corresponding device driver. A C library is responsible for implementing system calls. Usually it is GNU C Library (glibc). However, GNU C Library is not compatible for embedded Linux due to it is a large library. To quote from Ulrich Drepper, the maintainer of GNU libc: "...glibc is not the right thing for [an embedded OS]. It is designed as a native library (as opposed to embedded). Many functions (e.g., printf) contain functionality which is not wanted in embedded systems." 24 May 1999[3]. Thus, another C library, µClibc is introduced as shown in Figure 4. µ means micro while C stands for controller. And µClibc is the short of "the microcontroller C library". As the name it indicates, the µClibc is intended to support embedded systems that provide as much functionality as possible in a small amount of space.

III. SYSTEM IMPLEMENTATION

In this paper, the usability, license issues, pros and cons of open source IP based embedded system with Linux are investigated via a case study. The case is to develop an application runs on Linux that reads acceleration raw data from an I2C compass sensor-LSM303DLM. The hardware platform of this case is ORPSOC with I2C controller enabled.

The work flow for implementing this case is shown as following:

1. Set up the development environment includes Xilinx ISE for hardware synthesis and GNU tool-chain for software compilation.
2. Generate FPGA configuration bit stream of ORPSOC with I2C controller enabled.
3. Install u-boot.
4. Setup TFTP server and NFS server.
5. Build Linux image and install it through TFTP protocol.
6. Load the application software via NFS protocol to Linux to test the usability of whole system.

If the application is working, then the usability is proved. License issue is surveyed with interpreting license of each component involved in this thesis. And then check out if the usage of each component offends the rules. Pros and cons are obtained by comparing with the previous knowledge of conventional embedded system.

A. Prototype of open source IP based embedded system with Linux

1) The whole system is implemented on Atlys FPGA board including a Xilinx Spartan-6 FPGA chip[7].
2) The USB to JTAG interface is used to program SPI FLASH. The SPI FLASH image contains ORPSOC bit stream for configuring FPGA and binary image of the program that will be executed. In this thesis, the program is u-boot, which is a bootloader for further Linux installation.
3) The interaction interface between user and the board is via UART console.
4) Another host computer acts as TFTP and NFS server. The TFTP server is for transferring Linux image to Atlys board and NFS server is for loading applications to Linux.
5) I2C slave device is a compass sensor, LSM303DLM.

B. Generate ORPSOC with I2C controller enabled

In the source code of ORPSOC, there is a specific folder contains RTL source code and makefile scripts for Atlys board. The I2C controller is a soft IP core that should be added into ORPSOC by connecting it to Wishbone interface. However, it has been connected already and disabled as default in a top-define file "orpsoc-defines.v", which is for configuring functions of ORPSOC. Thus, uncommenting the code "`define I2C0" will enable the I2C controller in ORPSOC[8]. Then the next step is assigning the signals of I2C controller to PMOD socket developed by Digilent for peripheral connection[4]. The I2C interface requires 4 wires to communicate properly; they are Vcc, GND, clock line SCL and data line SDA. In ORPSOC, the names of SCL and SDA signals are "i2c0_scl_io" and "i2c0_sda_io" gained from top-design file "orpsoc_top.v". Figure 5 shows the PMOD interface on Atlys board and signal assignment[4].

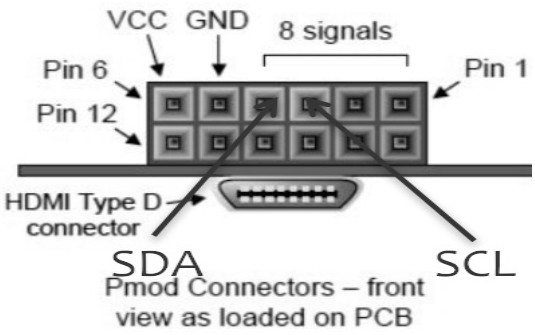


Figure 5: PMOD and I2C signal assignment.

SCL and SDA signals are assigned to Pin3 and Pin4 by editing user constraint file (ucf). The carrier board of I2C slave LSM303DLM has a level-shifter circuit that makes it compatible for PMOD 3.3V system. At last, the hardware platform ORPSOC is generated with the help of makefile scripts[5]. A bit stream for configuring FPGA can be achieved after the synthesis, mapping, place and route steps are finished. Figure 6 shows the block diagram of ORPSOC. The blue blocks are default settings, and the red I2C controller is the one enabled in top-define file.

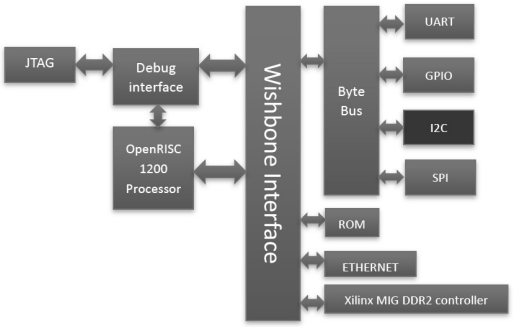


Figure 6: Block diagram of ORPSOC with I2C controller enabled

C. U-boot

Generally, GDB (GNU Project Debugger) is taking the responsibility of loading and debugging the software. However, the Xinlinx USB Platform Cable couldn't be detected by GDB. Thus, an alternative should be found to load the software onto the platform. The approach is u-boot. U-boot refers to universal bootloader, is a primary and powerful bootloader used in embedded systems that eases the procedure of loading Linux image or other application images. The u-boot can only fix the problem of loading software, whereas it can't be a perfect replacement of GDB because software can hardly debug with u-boot. The installation of u-boot is to generate a SPI FLASH image that contains the FPGA configuration bit stream and the u-boot binary image.

In ORPSOC, there is a ROM module written in Verilog language, which acts like a read-only memory. The ROM includes a bootloader program to load the software that is located in SPI flash. Therefore, the ROM and the bootloader are combined to be a bootrom. After the Atlys board is powered on, the FPGA will be configured and the u-boot is copied from SPI flash to external DDR2 RAM to start execution with the help of bootrom[6]. In addition, a block of 4 bytes should be added to the head of u-boot binary image that informs the bootloader in ROM how many bytes should be copied to RAM. In order to load Linux image through TFTP (Trivial File Transfer Protocol), the network parameters of u-boot and host computer should be set properly to build a LAN (Local area network) as shown in Table 1. The Ethernet MAC (Media Access Control) address can be random as long as it is valid if the system is only for experimental use. Moreover, the network speed of host computer should be configured to 100Mbps or below it because 100Mbps is the maximum connection speed supported by Ethernet MAC IP core in ORPSOC[4].

Table 1: Network parameters of u-boot and host computer

	U-boot	**Host computer**
Ethernet MAC address	00:12:34:56:78:9a	Assigned by manufacture
Local IP address	192.168.2.6	192.168.2.2
Network mask	255.255.255.0	255.255.255.0
Gateway IP address	192.168.2.1	192.168.2.1
Server IP address	192.168.2.2	-

D. TFTP and NFS server

TFTP (Trivial File Transfer Protocol) is a simple file transfer protocol without any authentication. Because of its simplicity, it is implemented with less consumption of memory. TFTP server helps transfer Linux images from host computer to the RAM of Atlys board.

The setup procedure of TFTP server is shown below:

1. Install the required packages: tftp, xinetd, tftpd
2. Create a folder for storing images that will be transferred to u-boot. And more important, the folder and the files in it should be configured as everyone can access it.
3. Create a file named "tftp" under "/etc/xinetd.d" path, which is the configuration file for tftp server. Start the service. NFS is defined as Network File System. A NFS server can be mounted on any device supports NFS protocol. Using NFS server during testing avoids copying applications to target board, thus the debugging procedure becomes convenience.

The steps for configuring NFS server is shown below:

1. Install the necessary components: portmap, nfs-kernel-server
2. Edit the file "/etc/exports" to add the entry will be shared.
3. The same as TFTP server, the shared folder and the files it contains should be configured accessible for everyone.
4. Start NFS server and portmap.

E. Building Linux kernel

Before building Linux kernel, a DTS (device tree) file should be created at first. When Linux is booting up, it is necessary for Linux to know what hardware resources the board has. Thus, the device tree plays this role. Device tree is a data structure that describes the hardware and it is passed to Linux kernel by the bootloader. Then the Linux kernel can configure the hardware and work appropriately. The device tree for Atlys board is derived from the one for OpenRISC simulator located in Linux kernel source code. Firstly, the clock frequency and memory size should be corrected to 50MHz and 128MB. And then the I2C controller node can be added according to an example of I2C controller description for device tree gainedfrom Linux's opencores I2C driver. Appendix C shows a complete device tree for this case. For Linux kernel, the I2C device interface support and I2C hardware driver for opencores are attached via "menuconfig", which is a tool for selecting the features of Linux kernel. Then the binary image of Linux kernel can be generated by the GNU tool-chain for Linux development. More than that, the image should be modified to a bootable image for u-boot with the tool "mkimage" provided by u-boot. U-boot refers to universal bootloader, is a primary and powerful bootloader used in embedded systems that eases the procedure of loading Linux image or other application images.

IV. RESULTS

A. Usability

The application reads sensor output successfully that proves the Open source IP based embedded system with Linux is usable for industry. Although the application in this thesis is just to read output from an I2C sensor, but its success demonstrates that the open source IP based embedded system with Linux is usable. With the help of various open source IP cores and Linux drivers, it is possible to develop more complex embedded system for industry usage.

B. Pros and cons

Table 2 is a comparison between open source IP based embedded system with Linux and conventional embedded systems. The open source IP based embedded system with Linux refers to the one whose hardware platform is ORPSOC having Linux installed.

Table 2: A comparison between open source IP based embedded system with Linux and conventional embedded systems

	Open IP based embedded system with Linux	Conventional embedded system without Linux	Conventional embedded system with Linux
Product cost	High	Low	Low
Development environment	Basic-supported	Well-supported	Well-supported
Flexibility and scalability	High	-	Only Linux has scalability
Software Portability	High	Low	High
Software development difficulty	Low (except the case if hardware is not supported by Linux driver)	Normally is high, depends on the vendor's support	Low (except the case if hardware is not supported by Linux driver)
Reusability	High	-	-

The conventional embedded system refers to the one whose hardware platform is based on IC chips with fixed resources, such as AVR32 and ARM architecture. Developers can choose to install Linux on it or not.Obviously, the pros of open source IP based embedded system with Linux are high flexibility and scalability, high software portability, low software development difficulty and high reusability. Whereas the cons are high product cost due to the higher price of FPGA chips, basic-supported development environment and more difficult software development if Linux driver doesn't support the hardware.

V. CONCLUSIONS

In this paper, a prototype of open source IP based embedded system with Linux is presented. A successful application reading the output of acceleration raw data from an I2C compass sensor-LSM303DLM is developed as well, which proves the usability of such an embedded system. Moreover, the investigation of licenses indicates the implementation doesn't offend the rules of GPL and LGPL and developing proprietary software on the system is valid. Hence, companies are allowed to develop and sell the embedded systems without providing their source code to users or public. Comparing to conventional embedded systems, although the open source IP based embedded system with Linux has some disadvantages such as: high product cost, only basic-supported development environment, and more difficult software development if Linux driver doesn't support the hardware. However, it has: high flexibility and scalability, high software portability, low software development difficulty and high reusability.

Open source IP based embedded system with Linux is hard for beginners to start with and the application development for Linux takes time to learn. However, once the developer handles them well, the time to develop an embedded system will be decreased significantly because the lower software development difficulty of Linux user space. Moreover, when the function of such an embedded system is changed, a total re-design of the embedded system can be avoided due to it is flexible, scalable and reusable. In addition to function change, the work of function migration can be reduced because of the high software portability and reusability. Hence, the open source IP based embedded system with Linux is more suitable in industrial usage.

REFERENCES

[1] Jason G. Tong, Ian D. L. Anderson and Mohammed A. S. Khalid, "Soft-Core Processors for Embedded Systems", 2006 International Conference on Microelectronics, Dec. 2006, pp.170-173

[2] Younghoon Bin, Kwangmyong Kang, Hyungjun Kim, Hongkyun Jung, Kwangki Ryoo, "The Development of SoC Platform for Embedded System Applications", 2007 International Conference on Convergence Information Technology, Nov. 2007, pp.2286-2291

[3] Lihong Lian, Xiaochao Li, Fen Xiao, Donghui Guo, "Design and implementation of a debugging system for OpenRISC processor", 2008 2nd International Conference on Anti-counterfeiting, Security and Identification, Aug. 2008, pp.368-371.

[4] Jiesheng Wei, Ling Wang, Feng Wu, Yibo Chen, Long Ju, "Design and implementation of wireless sensor node based on open core", 2009 IEEE Youth Conference on Informtion, Computing and Telecommunication, Sept. 2009, pp.102-105.

[5] Faroudja, A. ; Izeboudjen, N. ; Titri, S. ; Sahli, L. ; Louiz, F. ; Lazib, D, "Hardware/Software development of a System on Chip platform for VoIP application", Microelectronics, Dec. 2009, pp.62-65

[6] Mehdizadeh, N. ; Shokrolah-Shirazi, M. ; Miremadi, S.G, "Analyzing fault effects in the 32-bit OpenRISC 1200 microprocessor", 2008 Third International Conference on Availability, Reliability and Security, March 2008, pp.648-652

[7] Castillo, J. ; Huerta, P. ; Lopez, V. ; Martinez, J.I, "A secure self-reconfiguring architecture based on open-source hardware", 2005 International Conference on Reconfigurable Computing and FPGAs, Sept. 2005, pp.7 pp.-10

[8] Younjin Jung, Ok Kim, Byoungyup Lee, Hongkyun Jung, Kwangki Ryoo, "SoC platform design with multi-channel bus architecture", 2008 International SoC Design Conference, Nov. 2008, Vol.03, pp.III-48-III-49

C Bounded Model Checker to Detect Unspecified Expression in FreeRTOS

P.Venkata Gopi kumar
Department of EIE, VNR VJIET, Hyderabad, India
Email: gopikumar_pv@yahoo.com

Abstract—**The Embedded systems are widely used in most electrical devices. They are often complex and safety –critical. Therefore their reliability is significantly important. Among many techniques to verify a system, model checking models a system into temporal logic and can be used to assert a desired property on it. CBMC is a bounded model checker for ANSI-C and C++ programs. In this paper , it is extended the CBMC tool to check and automatically detect a C/C++ code containing a form of un specified behaviors, like function calla with arguments that exhibits side effects which might be easily un noticed by the programmers. In addition, the code can be configured properly to be used for Arm Cortex micro softwares..**

Index Terms—**Embedded system, CBMC, RTOS, BMC.**

I. INTRODUCTION

The majority of computer devices are embedded systems. These days cell phones, cameras, home appliances, robots, industrial machines, traffic lights, trains, airplanes, and many other devices mainly contain an integration of computer systems. Embedded systems are often complex and safety-critical. As both their hardware and software complexity are significantly increasing, reliability moves into the center of attention and needs to be tested properly.

Testing could be done in different stages, while producing software and it is sometimes as complex and time consuming as developing the software. Therefore, it is more beneficial to find bugs at an early stage in software development and provide valuable feedback for developers, in order to find and fix such problems prior to building up the next modules or even next release. Sometimes bugs are due to a bad usage of a documented library or API, because of not reading the whole manuals or misunderstanding them. Another case could be not only a programmer's mistakes but also by reason of using complicated programming languages, like C/C++. In fact, while C/C++ are the most widely used languages for developing such systems, they are counted as highly prone to errors. These errors might lead to very serious consequences, including unpredictable and inconsistent program behaviors, run-time errors and even system crashes. Consequently effective detection of such errors is necessary.

Many embedded system applications use a special operating system called Real Time Operating System (RTOS). FreeRTOS is an open source RTOS that is used for embedded platforms such as ARM, Cortex-M3, AVR and STM32. It is written mostly in C and offers a small and simple real time operating system. It provides one solution for many different architectures and development tools and it is known to be reliable. In this paper, we use FreeRTOS as a processor of targeting program which might contain unspecified behavior.

For this purpose, we prepared a minimalist or simplified model of the FreeRTOS API in form of a C library. For achieving error detection goal, there are variable verification techniques. Using formal methods are well-known, which mathematically specify and verify these systems. They give us a proper understanding of a system and reveal inconsistencies, ambiguities, and incompleteness that might otherwise go undetected [1].

One of the most widely used formal methods is model checking, a technique that relies on building a finite model of a system and checking if a desired property holds in that model. *Bounded Model Checking* (BMC) techniques are able to efficiently and statically detect the possibility of run time exceptions in low-level imperative code, i.e., due to erroneous use of pointers, arithmetic overflows, or incorrect use of APIs. One of the most successful tools for automatic verification that implements the bounded model checking (BMC) technique, is the C Bounded Model Checker (CBMC) used for ANSI-C/C++ programs. There is a class of defects that is detected by this CBMC, while many other verification tools have not unnoticed yet. For instance, among many features, we emphasize more on its ability to check array bounds even with dynamic size, pointer safety during conversion of pointers from and to integers and user-specified assertions; moreover it models integer arithmetic accurately, and is able to reason about machine-level artifacts such as integer overflow. In CBMC, any sort of checking appears as a specification that comes to a boolean formula, which is then checked for satisfy ability by using an efficient SAT procedure. As a result, either a counterexample is extracted from the output of the SAT procedure, in the case that the formula is satisfiable, or if the formula is not satisfiable, the program can be unwound more to determine if a longer counterexample exists [2].

We extended the *CBMC* tool to check and automatically detect C/C++ code containing one form of unspecified behavior in the C/C++ standard which might go easily unnoticed by the programmers. According to the C/C++ standard, the order in which the arguments to a function are evaluated is unspecified and it depends on many factors like the argument type, the architecture, the platform and the compiler. The standard dictates that a C/C++ implementation may choose the order in which the function arguments are evaluated. It is the programmer's task to take care of them and make sure that the program

does not depend on the order of evaluation. However, there is a warning flag in C/C++ compiler like GNU, *-Wsequence-point,* which warns about code that may have unspecified semantics because of violations of sequence point rules in the C and C++ standards. However, the current approach is suboptimal and many complicated cases are not diagnosed by this option. For instance, through this flag, the side effect among the array indexes are not evaluated precisely. If two indexes are expressions that might get same value at some point in the code, the flag is not able to detect this case.

Eventually, it is provide capability for proposed CBMC extension to be run on applications written in FreeRTOS, added an option to the CBMC front-end to verify if a given C/C++ code contains no such kind of side effects in arguments of each function and warn the programmer if there is any evaluation order dependency in the code..

This paper presents a study to develop a method and an automated tool for automatic detection of software defects. The target programs are written using the FreeRTOS real-time operating system, compiled by the ARM Micro-controller Development Kit (MDK-ARM) and executed on ARM Cortex micro-controllers. The starting point of this work was the existing bounded model checker CBMC. We extended *CBMC* to be able to model check C code for ARM Cortex micro-controllers and automatically detects software defects in FreeRTOS softwares such as general C faults like function calls with arguments that exhibit side effects. It covers any kind of expressions containing variables, structures, classes, arrays and arithmetic operators over them. Moreover, we consider sequence points which force the compiler to evaluate the expressions in predefined order such as ‖ , &&, ?: and comma. For evaluating the result, we compared our modified CBMC with the original CBMC and Coverity verification tool[15], [16] and GNU Compiler Collection (GCC) using *-Wsequence-point* flag. The result shows that the extended CBMC has the ability to check more unspecified behavior than the other three tools. Next, author prepared a minimalist model of the FreeRTOS API in the form of a C library, in order to support detection of defects that might happen in FreeRTOS software specific to the MDK-ARM compiler and explained why it is essential.

II. EXISTING MODEL CHECKERS

In this section chapter, we briefly introduce important concepts and tools that are used in this work. Verification is a procedure of evaluating if a system meets a specification or imposed requirements. To verify a system, among many formal methods, model-checking and theorem proving are well-known. They are mainly used to analyze the system based on its specification for certain properties.

Model checking requires building a finite model of a system, It checks whether a desired property holds in that model. There are several ways to model check C code such as Bounded Model Checking (BMC), model checking with predicate abstraction using a theorem prover, model checking with predicate abstraction using a SAT solver and translation of the C code into a model of an existing standard model checker [3]. The common property in all these techniques is an abstracted program with a finite state space that is gained from transformation of the system. Finiteness is required because the model checking algorithm should go through all states.

Bounded Model Checking, as the name suggests, does this transformation by unwinding possibly infinite constructs a finite number of times, for example, it executes *while* loops *n* times, where *n* is a limiting upper bound. A tool that implements bounded C model checking is *CBMC*. It is able to find a suitable *n* in most cases. However, if it does not succeed, there is a possibility for users to provide their own upper bound to be used by *CBMC*. In such a case, CBMC cannot guarantee that the user-provided upper bound is long enough to not miss any errors and that no longer counter-example is available. This is the case, where *CMBC* can only find errors and not prove correctness [3]. The advantage of model checking over theorem proving is that model checking can be used to check if a system is completely specified or to verify modules or partial specifications. It is completely automatic and fast and contributes useful information of system's correctness. The model checker will either terminate with answer true indicating that the model satisfies the specification or give a counterexample execution that shows why the specification is not satisfied, which can be useful while debugging is shown in Figure 1 [4].

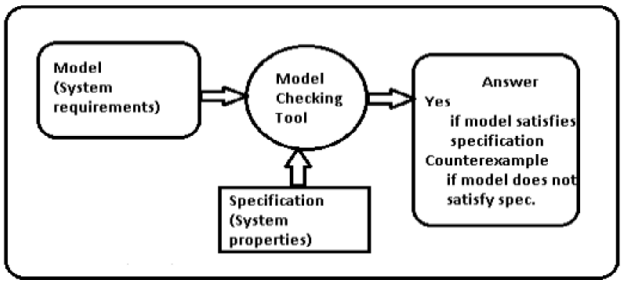


Figure 1: The model-checking approach

In theorem proving, a system and all desired properties are revealed mathematically in formulas. This is given by a formal system, which defines a set of axioms and a set of inference rules. Then all properties that should be held by the system are being proved from the axioms of the system by applying the inference rules. It is essentially, a process of proving a property from the axioms of the system. In contrast to model checking, theorem proving concerns infinite state spaces and proves these domains by structural induction techniques. Theorem proving mainly requires interaction with a human and humans might yield invaluable vision into the system and properties for being proved and it makes the process slow and sometimes error prone [1].

III. C BOUNDED MODEL CHECKER (CBMC)

CBMC is an open source model checker that uses bounded model checking technique to verify C or C++ programs. C/C++ files names are given to it as command

line arguments. Similar to other compilers, it integrates all definitions and functions from each file but instead of making the binary code, it produces a goto-program of the program. The goto-programs are simplified C/C++ programs, which contain program's information such as variable's data type, any type casting and library functions, in a structured way and are represented in the form of Control Flow Graphs (CFG). In goto-programs, each variable is assigned once and it is done by renaming in each case, this transformation is called Static Single Assignment (SSA).

In next step, a CNF is generated from this intermediate symbolic code and is passed to a SAT solver. SAT solver checks this equation's validness and it gives a counter-example trace when it fails. This shows that a bug is found in the program [2]. Considering the real time behavior of embedded systems, loop constructs are limited in number of iteration. CBMC verifies such finite upper run time bounds by unwinding all loops and checks if enough number of iterations are set in order to prove the absence of errors[2]. CBMC also provides set of keywords, which can be used to aide CBMC with more information about the program. These keywords can be used for program instrumentation. The program instrumentation is a procedure to verify some properties of the code. *CPROVER_assert(expr)* and *CPROVER_assume(expr)* macros are examples of these keywords. The former can be used to check any condition (*expr)* with the same logic for assertions in the usual ANSI-C expression logic. When CBMC encounters this keyword, it tries to generate a formula to check assertion failure. The generated formula is verified using SATsolvers. If the formula is satisfiable then assertion fails and CBMC generates error and produces counter-example showing possible trace of error. The latter macro, *CPROVER_assume(expr),* is used to restrict non-deterministic choices made by the program and it reduces the number of program traces that are considered and allows assume-guarantee reasoning [5].

CBMC also supports pointers, arrays, structures, floating point operations and function pointers. There are other tools like BLAST [6] and Extended Static Checker for Java (ESC/Java)[7]. BLAST is a software model checker for C programs. Like CBMC, it checks that software satisfies behavioral properties of its interfaces and it uses counterexample-driven automatic abstraction refinement to construct an abstract model, which is model checked for safety properties. However, the advantage of CBMC over BLAST is that CBMC can also be used to verify consistency of hardware designs with a functional specification (written as C program). It can verify modules, and not only whole programs and it treats recursive functions and has GUI. ESC/Java tool also attempts to find common run-time errors at compile time but in Java programs. It is based on simplify theorem prover using SAT checking and translates code to SSA, and then into verification conditions. ESC/Java supports assume-guarantee reasoning that are on methods and method calls, whereas in CBMC assume-guarantee statements can appear in any place in the program.

IV. FREE RTOS OPERATING SYSTEM

Real time systems often run on special operating systems. A Real Time Operating System (RTOS) provides facilities to programmers such as process execution, predictability, data structures, and mechanisms for inter-process communication. FreeRTOS is used to develop real time systems for embedded devices. FreeRTOS is designed to be small and simple. The kernel itself consists of few C files. To make the code readable, easy to port, and maintainable. It is written mostly in C, but there are a few assembly functions included where needed (mostly in architecture specific scheduler routines). FreeRTOS provides methods for multiple threads or tasks, mutexes, semaphores and software timers [8]. The fast execution, low overhead, configurable scheduler, co-routine supports, trace support
and very small memory footprint are key features of FreeRTOS.

In C and C++ standards, the order of evaluating expressions is expressed by concept of sequence points. A sequence point shows which part of the expression is executed before and which one after it. Therefore, a partial ordering occurs between executions of different sides. For instance, sequence points could be after the first operand of operators &&, ‖ and ?:, in a function call after evaluation of its arguments but before executing the
function body and in many other specific cases.

V. UNSPECIFIED SIDE EFFECTS

In this section, it is describe how evaluation of function arguments could be seen as one of the common defects in C/C++. Both expressions give evaluation and defects in formal semantic. In the last part of this chapter, it is explained our algorithm used and show how these side effects in function arguments are detected. As it mentioned earlier, bugs could occur due to bad usage of the documented rules of programming languages. This is quite common in C/C++ programs. Sometimes programmers forget to check if their codes are specified by the standard. More specially if the code has portable behavior and they can count on it. Our focus is on how this could be issued in evaluation of function arguments.

A. Undefined Behavior

Behavior, due to use of a non-portable, erroneous data or program construct, where the standard imposes no requirements for them. An example of undefined behavior is the behavior on integer overflow. Behavior, where each implementation documents how the choice is made and the language provides a documentation describing its characteristics and behavior. An example of implementation-defined behavior is size of integer where the implementation must have only one definition for every place in the program.

B. Unspecified Behavior

Behavior, where a set of allowable possibilities is defined but it is not deterministic. The standard enforces no further requirements and the implementation is not required to document which option is chosen in any

occurrence. For example, the compiler can choose different possibilities in different places, where the cases could even happen in the same program. Moreover, from the C standard specification, we mark the following cases that are not specified in the language [11]: Use of an unspecified value, or other behavior where the International Standard provides two or more possibilities and imposes no further requirements on what is chosen in any instance. An example of unspecified behavior is the order in which the arguments to a function are evaluated (§3.4.4) The order in which sub-expressions are evaluated and the order in which side effects take place, except as specified for the function-call (), &&, ||, ?:, and comma operators (§6.5).

The order in which the function designator, arguments, and sub-expressions within the arguments are evaluated in a function call (§6.5.2.2) According to the C++ standard [9], the order in which the arguments to a function are evaluated is given as an example of unspecified behavior. In fact, it depends on many factors like the argument type, the called function's calling convention, the architecture and the compiler. On an x86, the Pascal evaluates arguments left to right, whereas in the C/C++ calling convention it is right to left. Therefore, programs, which run on multiple platforms should take the calling conventions into account to skip any surprises, side effects or crashes. The standard dictates that a C/C++ implementation may choose in which order, function arguments are evaluated. To be in the safe side, the program itself should not depend on the order of evaluation of side effects and shall not use parameters of a function in default argument expressions, even if they are not evaluated. By the following examples, we intend to clarify this common unspecified case according to
the standard. Consider the function *test:*

```
void test(int arg1, int arg2, …);
```

Assume that somewhere in the program there is a call like:

```
int i = 0;
test(i++, i, ...);
```

How or in which particular order, different environments evaluate the arguments, is so important that even this simple function call can behave differently from one to other. For instance, *test(1, 1, …)*, *test (1, 0, …)* or even *test (0, 0, …)* yeild possible results.

The second case is when arrays are involved; the index expressions come to center of attention.

```
int a[2] = {0, 1};
int i = 0;
test(a[i] ++, a[i], …);
```

But more interesting example is when we have different indexes of an array:

```
int a[2] = {0, 1};
int i = 0;
int j = 0;
test(a[i] ++, a[j], …);
```

In this case, from the syntax point of view, *a[i]* is not *a[j]*. However, they might point to the same location of memory when *i* and *j* hold same value. In addition, next example shows that the sequence point rule could effect these unspecified cases:

```
int i = 0;
test(..., i++ || i , ...);
```

The || operator is a sequence point and forces the compiler to evaluate its left and right operands in a specified order; then there is no unspecified behavior in this example. Therefore, it may be necessary to warn the user, if evaluation of arguments of any particular function lead to unspecified behavior due to expressions with possible side effects. However, the original CBMC allows all side effect operators with their respective semantic. Moreover, regarding the ordering of evaluation, CBMC uses a fixed ordering of evaluation for all operators. It believes, while such architecture dependent behavior is still valid in ANSI-C programs, showing these cases are not desirable [5].Furthermore, we saw these side effect warnings as a demand and added this option to CBMC front-end, to verify that a given C/C++ code contains no side effects in arguments of its functions. In the following section, we present a formalization of argument expression through precise description of the C/C++ language interface.

In this section, a formal semantics of expression evaluation is presented. The Structural Operational Semantics (SOS) is used in this project, which is a set of rules for giving a formal semantics of expression. It basically defines the behavior of a program in terms of behavior of its parts and provides a structural view on operational semantics; in my opinion this structure is easy to follow. An SOS rule is in the form of:

assumption , requirement
conclusion (*name*)

where the *assumption* is a pre-condition of an expression before its evaluation and *requirement* shows under which domain this assumption is hold.

Although this simplified grammar of expressions is not fully matched to C/C++ languages, it covers most main types of operators with clear syntax similarity to C/C++. For the sake of simplicity, the similar operators are skipped in this grammar but it is easily extendable without extra complexity.

```
int_expr ::= var_access |
int_expr bin_opr int_expr |
int_expr seqpoint_opr int_expr|
var_access ::= int_var |
int_var++ |
array_access |
array_access++
int_var ::= x
array_access ::= a[int_expr]
bin_opr ::= + | - | * | / | = = | <
seqpoint_opr ::= || | &&
```

C) Algorithm of Evaluation Order Side Effect

There are several constraints on how to evaluate expressions in C/C++ language standard. As mentioned before, the most significant one is that “between the previous and next sequence point an object shall have its stored value modified at most once by the evaluation of an expression. Furthermore, the prior value shall be accessed only to determine the value to be stored” [ISO90, x6.3]. Violation of this constraint might result in unspecified

behaviors. In this part, we present our algorithm for checking it and we explain how we determine the existence of the side effect in evaluation order of the arguments to any function.

The action of side effects happen by changing the memory. Therefore, it is important that while evaluating certain expression, any pairs of read and write over certain memory location are seen as a potential side effect. Principally, operators like assignments, increment or decrement are counted as write operators. For instance, in the following function's arguments, there are a few read and write pairs. Your goal is to simulate the usual appearance of papers in a Journal Publication of the CVR College. We are requesting that you follow these guidelines as closely as possible.

VI. IMPLEMENTATION

This section shows briefly the modifications made to CBMC tool to be able to find possible unspecified behaviors in a given source program.

A. CBMC

The argument side effect checking, described in chapter 4, is implemented using C++ programming language. The source code is checked out from subversion (SVN) repository http://www.cprover.org/svn/cbmc/

We used the *trunk* version for windows in this thesis. In order to reduce the amount of work required to set up a Visual Studio project for CBMC and the associated tools, a script is used which automates this process. The script is available in the CBMC SVN *trunk* in the directory "scripts" and is called "generate_vcxproj". It could be configured by following command in a bash shell, e.g., provided by cygwin. ./generate_vcxprojThe command reads the Makefiles and automatically generates project files for cbmc, gotocc and goto-instrument, and we can access them through Visual Studio. The project files come with filter definitions that order the source files according to the (top-level) sub directories they are in.

Note that the flex and bison tools and the irep_id conversion tool still need to be run manually as mentioned in the compiling hint document.

This project file is helpful for debugging and building with MSBuild. For windows platform, CBMC still requires the pre-processor cl.exe, which is part of Visual Studio and the path to cl.exe must be part of the PATH environment variable of your system. The *trunk* is structured in a similar fashion to a compiler. It contains language specific frontends with limited syntactic analysis, intermediate format and a back-end tool for processing this format. Like a compiler, it takes the names of .c/.cpp files as command line arguments, then it translates the program and merges the function definitions from the various .c/.cpp files, just like a linker. But instead of producing a binary for execution, it performs symbolicsimulation of the program[13]. Here, we outline the trunk project but only the important directories with files that get modified, for the sake of clarity.

/trunk

/src

All source codes are located in this directory and they are separated into different sub directories, such as, /analyses, /cbmc, /goto-programs, etc.

/goto-programs

Contains the transformation program of the source code to an intermediate representation of C/C++ which is language independent. All converting methods are located here, and our new support is mostly added as a goto-program.

/cbmc

The first full application is this directory. Here, the front ends (ansi-c, cpp, gotoprogram or others) are used to create a goto-program, goto-symex to unwind the loops the given number of times and produce and equation system It then uses solvers to find a counter-example.

/goto-cc

It is a compiler replacement that just converts C/C++ programs to goto-binaries. It is supposed to be dropped into an existing build procedure in place of the compiler Thus, it emulates flags that would affect the semantics of the code produced. Which set of flags are emulated depends on the naming of the gotocc/ binary. If it is called goto-cc then it emulates GCC flags, goto-arm cc emulates the ARM compiler, goto-cl emulates VCC and goto-cw emulates the Code Warrior compiler. The output of this tool can then be used with cbmc[13].

/goto-instrument

The *goto-instrument* is the top level control for the program. It could be used as a skeleton of new project. This directory contains a number of tools that are used in a goto-program. One can either modify it or perform some analysis. Here the command line is parsed to see which option is desired by user. It supports the following checks:

--no-assertions ignores user assertions
--bounds-check adds array bounds checks
--div-by-zero-check adds division by zero checks
--pointer-check adds pointer checks
--*arguments-check** adds argument order checks
* not available in original CBMC

/analyses

It makes a list of all checks that should be analyzed (e.g. options taken as arguments by command line parsing).

/doc

The html and pdf versions of the source code documentation explaining the above directories more detailed [13]. We also need a SAT solver (in source). MiniSat2 is recommended by CBMC and it could be downloaded from: http://minisat.se/downloads/minisat-2.2.0.tar.gz

B. CBMC Extension

To design the argument-checker that was discussed in Chapter 4, we add a module to *goto_programs* directory. Knowing some of the basic concepts might be useful here, such as, each function is a list of instructions, each of which has a type (one of 18 kinds of instructions), a code expression, a guard expression and potentially some targets for the next instruction. Our module checks each expression while it is being converted to an intermediate format referred to as *goto-binaries* or *goto-programs*. In

conversion level,CBMC has a technique to adjust the code to standard definition in some special cases and prevent some side effects by cleaning expressions like && ‖ ?: comma (control dependency), ++ --, compound assignments, object constructors like arrays, string constants, structures and function calls. It actually rewrites the expression in a way that the standard specified. However, as we like to check expressions with more sensitivity, we need to do it before any cleaning to ensure nothing is missed or basically converted. In this regard, we make a list of identifiers of arguments list for each function. Each identifier represents a variable used in arguments in a function.

VII. EVALUATION AND CASE STUDY

This section summarizes the result of model-checking performed some codes containing undefined behavior in their argument list of functions. Experimented with the same code with desired dependency among a function's arguments through Coverity 7.0, GCC 4.8.2 and our modified CBMC.

The case code contains two types of dependency among arguments For clarity, injected them in separate functions.

```
int a = 0;
int c = a;
size_t order[3] = {1, 2, 3};
get_order(order[a], order[c]++);
get_order(order[a], order[a++]);
```

After testing the code by the mentioned tools, it is observed that all three tools are able to detect some sort but not all kinds of evaluation order dependencies in both functions arguments. In this test code, line 17 is reported in all three compiling ways as evaluation order violation due to a pair of *read* and *write* operations over variable *a.* Similarly, experienced more dependencies in variables of expressions with no array memories and all these tools found them successfully. However, in different type of dependencies the result was not the same; For example, in this code, in line 16, when indexes *a* and *c* of array *order* is read and written respectively, Coverity and GCC are not able to check whether these indexes hold same value and if the same location of memory is going to be processed or not.

In contrast, the modified CBMC is able to detect it. Figure 3 shows that modified CBMC found this possible violation. Moreover, CBMC creates a counter example trace which is a program trace that ends in a state which violates the property *(a==c). andFigure shows the GCC 4.8.2.*

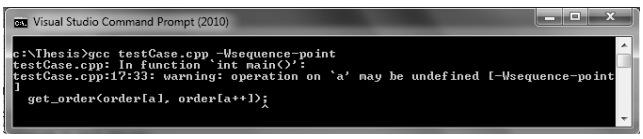


Figure 2: GCC 4.8.2

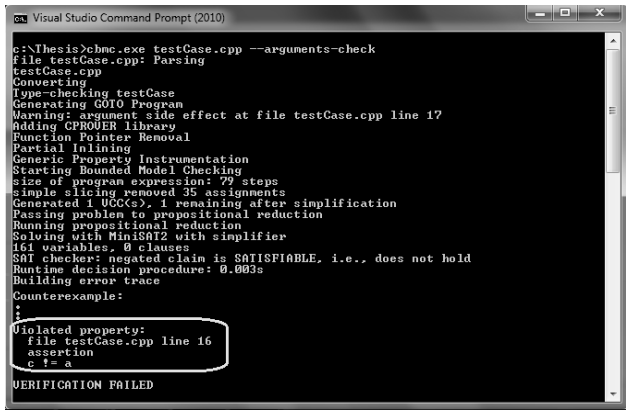


Figure 3: Modified CBMC

VIII. CONCLUSIONS

In this work, author extended the CBMC to verify real-time programs run on FreeRTOS operating system and MDK-ARM firmware and specially found some possible unspecified behaviors. The targeting program might contain unspecified behaviors, such as, when evaluation order of arguments to a function are not defined by the standard as it depends on many factors like the argument type, the called function's calling convention, the architecture and the compiler. This dependency among expressions could lead to non deterministic behavior of a system and causes serious issues. For this purpose, a method is prepared to detect such an unspecified behavior by extending available tool named CBMC and we equipped a FreeRTOS API to be able to utilize this modification. The CBMC tool was easy to extend and working with it was simple and instructive as it is an open source tool and supported by valuable tutorial and full documentation. Its good reputation and being a notable tool for testing C/C++ programs motivated us to add more supports into it.

In conclusion, it is observed that, the proposed tool worked well at detecting a wide range of different dependencies in expressions of a function's arguments, including direct access to memory locations or through array indexes. As future work, it could support expressions containing such dependencies, when the pointers of same memory location are involved. It is very similar to cases with arrays that are already included. Further, the current code checks these unspecified behaviors specifically among arguments of any function in a program. The same method can be extendable to check them in any expressions in the whole program.

REFERENCES

[1] E. Clarke and J.M. Wing. *Formal methods: State of the art and future directions.* ACM Computing Surveys (CSUR), 28(4):626-643, 1996.

[2] D. Kroening. The cbmc homepage. http://www.cprover.org/cprovermanual/ Introduction.shtml, April 2013.

[3] B. Schlich and S. Kowalewski. *Model checking C source code for embedded systems.* International Journal on

Software Tools for Technology Transfer. Volume 11, pp 187-202, July 2009.

[4] E. Clarke. *Model checking*. In Foundations of Software Technology and Theoretical Computer Science, pages 54-56. Springer, 1997.

[5] E. Clarke and D. Kroening. Ansi-C bounded model checker user manual. Technical report, Technical report, School of Computer Science, Carnegie Mellon University, 2006.

[6] B. Dirk, H. Thomas A., J. Ranjit and M. Rupak. *The Software Model Checker Blast*. International Journal on Software Tools for Technology Transfer **9** (5-6): 505–525, 2007.

[7] C. Flanagan, K.R.M. Leino, M. Lillibridge, G. Nelson, J. B. Saxe and R. Stata. *Extended static checking for Java.* In Proceedings of the Conference on Programming Language Design and Implementation, pages 234--245, 2002.

[8] R. Barry. *FreeRTOS Reference Manual* - API Functions and Configuration Options, Real Time Engineers Limited, 2009.

[9] P. Becker, *Working Draft, Standard for Programming Language C++*, http://www.openstd. org/jtc1/sc22/wg21/docs/papers/2005/n1905.pdf , 2013.

[10] *Warningoptions*, http://gcc.gnu.org/onlinedocs/gcc/Warning-Options.html, 2013.

[11] Bruno R. Preiss (1998). *Expression Trees*. Retrieved December 20, 2010.

[12] R. Barry. *Using the FreeRTOS Real Time Kernel - a Practical Guide*, generic CORTEX
M3 edition. 37

[13] M. Brain, M. Tautschnig. *Beginner's Guide to CPROVER.* March 2014.

[14] Vijay D'Silva, Daniel Kroening, Georg Weissenbacher. *A Survey of Automated Techniques for Formal Software Verification.* IEEE Transactions on Computer-Aided
Design of Integrated Circuits and Systems (TCAD), vol. 27, no. 7, pp. 1165–1178, July 2008.

[15] The *Coverity® 6.6 Deployment Guide,* 2003-2013 Coverity, Inc.

[16] *Coverity 7.0 Checker Reference*, 2004-2014 Coverity, Inc.38

Design and Implementation of Instrumentation Amplifier for EEG in 180nm CMOS technology

O. Venkata Krishna[1], B. Janardhana Rao[2]
[1]Department of EIE, CVR College of Engineering, Hyderabad,India
Email: venkatakrishna.odugu@gmail.com
[2]Department of ECE, CVR College of Engineering, Hyderabad,India
Email: janardhan.bitra@gmail.com

***Abstract*— The EEG is Electro Encephalo Gram that which gives the electrical activity of the brain. Analysis of the electrical activity is very important to diagnose neurological problems. But, the problem generally involved is that the EEG signals have very low amplitude and these very low amplitude signals cannot drive displays for further analysis. So, we require to use amplifiers. Instrumentation amplifier is widely used for such applications because they give high gain. For designing the instrumentation amplifier, we have to first design the operational amplifier and this is further used to design the instrumentation amplifier. The 180nm CMOS technology has been used for the Implementation of this amplifier. It gives a hiah gain of 111 dB an low power dissipation of 16mW.**

***Index terms*—EEG, Instrumentation amplifier, Op-Amp, CMOs.**

I. INTRODUCTION

Measurement of brain signals is necessary to diagnose neurological disorders and other brain disorders. To diagnose, the brain signals are taken during sleep and other conditions of the patient. With the obtained results, the doctors conclude the disease and continue with the treatment. The Electrodes are placed on the scalp of the patient and these electrodes give an electrical output whose amplitude is in terms of micro volts. A protection circuit is used to avoid damage to the setup. These electrical signals are further amplified by an instrumentation amplifier and filtered and sent to a display device. The EEG is also vital for human machine interface that can be used for non-verbal communication. Electrodes are placed in specific locations of the scalp.

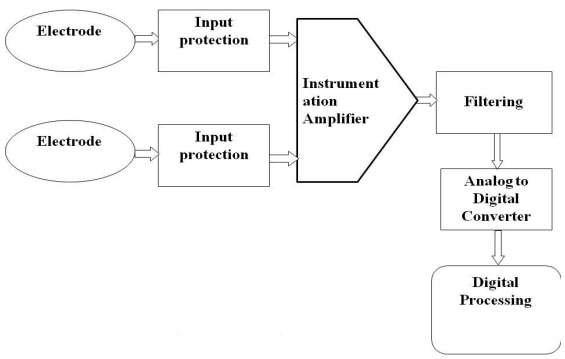


Figure.1: EEG signal chain.

When we look at the block diagram in Figure.1 , we understand the complete operation of the instrument. The electrodes collect the signal from the brain and given to amplifier and then filtered. The amplification occurs first because the brain waves have too low amplitude and frequency and it is very important for the signal to strengthen to separate artefacts and the EEG through filtering. There are many artefacts that cause distortions in the brain waves, they could be both technical and patient related. We can see even if a small part of the body is moved, even when they eye is moved, could be even due to sweating and even due to too much electrode jelly and low battery. We have filters like high pass, low pass, notch filters that allow only the brain waves. These filtered waves are then converted into digital signal and processed for further analysis.

From the above block diagram of an EEG it is evident that we require an amplifier for further analysis of the brain waves. It is also evident from the clinical applications of an EEG that measuring brain waves is vital to analyze the brain and conclude on diseases related to the brain. The instrumentation amplifier used has to be designed so as to meet the criteria of an EEG with following specifications.

Table 1. Specifications of Instrumenation Amplifier.

PARAMETER	VALUE
Input impedance	10Mohm
CMRR	>80dB at 50/60Hz
Noise	<2μV peak to peak
Bandwidth	0.1 to 95Hz, -3dB
Gain	3000
Impedance checking	Built in, 10Hz sine wave. Range 1kilohm to 50kilohms, 10%
Samples rate	200/400 samples/sec/channel, simultaneous sampling
Input signal range	2mV peak to peak full scale
DC input voltage range	±450mV max
Slew rate	10V/μs

II. DESIGN OF INSTRUMENTATION AMPLIFIERS

The instrumentation amplifier is similar to the differential amplifier that has input buffer amplifiers fitted to remove the need for extra circuitry required for impedance matching making these amplifiers useful for measurement and test equipment is shown in figure 2.

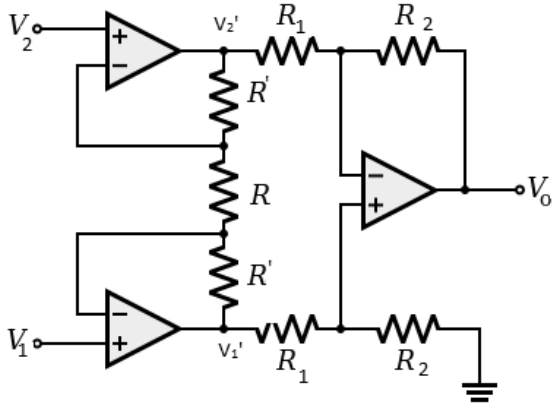


Figure 2: Instrumentation amplifier

The instrumentation amplifiers can be built with individual op-amps and precision resistors, but are also available in integrated circuit form from several manufacturers (including Texas Instruments, National Semiconductor, Analog Devices, Linear Technology and Maxim Integrated Products). An IC instrumentation amplifier typically contains closely matched laser-trimmed resistors, and therefore offers excellent common-mode rejection. Examples include AD8221, MAX4194, LT1167 and INA128.

Instrumentation amplifiers offer a unique combination of differential inputs, high input impedance, and excellent precision and noise specifications. Using both zero-drift and traditional topologies, Linear Technology's instrumentation amplifiers feature high precision, low drift and excellent PSRR and CMRR. Like all Linear Technology devices, our instrumentation amplifiers are unique in offering fully specified, tested and guaranteed performance for key parameters over the full operating temperature range, enabling high reliability designs.

Instrumentation amplifier is the front end component of every measuring instrument which improves the signal to noise ratio of the input electrical signal from the transducer. It uses the fact the noise is common to the both output terminals of a transducer across which the output is measured and sent to measuring instrument. Advantages of instrumentation amplifier High gain, the gain of the instrumentation amplifier can be varied by just varying resistors in input circuit without affecting the resistors in difference amplifier circuit, high CMMR, High input resistance.

To design an instrumentation amplifier an operational amplifier has to be designed first. We generally use a two stage operational amplifier since a single stage amplifier will have low gain. The proposed instrumentation amplifier was designed using 180nm CMOS technology. The technology provides six levels of metal (copper and aluminum) with Vth values of 0.48 V and -0.4 3 V for the NMOS and PMOS transistors, respectively. An extension to that technology provides optional passive devices for analog circuit design, including MIM (Metal-Insulator-Metal) and thick oxide MOS capacitors.

To understand the design of the instrumentation amplifier, we first consider a differential amplifier as shown in figure3.

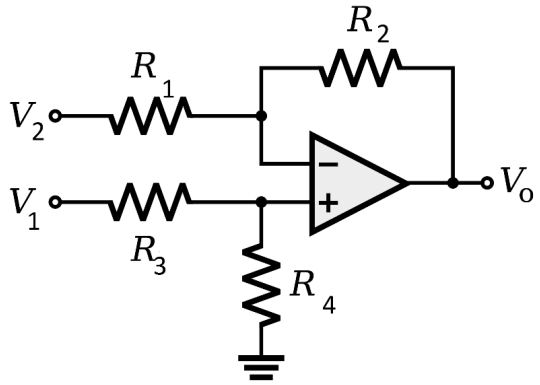


Figure 3: Differential amplifier

We use the super position principle and using that, when V_2 is applied alone:

$$V01 = -\frac{R2}{R1}.V2$$

When V_1 is applied alone:

$$V02 = \left(1+\frac{R2}{R1}\right).V' = \left(1+\frac{R2}{R1}\right).\left(\frac{R4}{R3+R4}\right).V1$$

Where,

$$V' = \frac{R4}{R3+R4}.V1$$

The output voltage V_o can be obtained as follows:

$$Vo = V01 + V02$$

$$Vo = -\frac{R2}{R1}.V2 + \left(1+\frac{R2}{R1}\right).\left(\frac{1}{1+\frac{R3}{R4}}\right).V1$$

Taking R2/R1 common, we get:

$$Vo = \frac{R2}{R1}.\left[-V2 + \left(\frac{R1}{R2}+1\right).\left(\frac{1}{1+\frac{R3}{R4}}\right).V1\right]$$

If R1/R2=R3/R4 then we have:

$$Vo = \frac{R2}{R1}\left[-V2 + \left[\left(\frac{R1}{R2}+1\right).\left(\frac{1}{1+\frac{R1}{R2}}\right).V1\right]\right]$$

Simplifying,

$$Vo = \frac{R2}{R1}.[-V2 + V1]$$

Rearranging the terms,

$$Vo = \frac{R2}{R1}.[V1 - V2]$$

To avoid the problems of input impedance we add buffer circuits to the differential amplifier and it appears as shown in Figure 2. It represents a three op-amp instrumentation amplifier where the first two operational amplifiers act as input buffers and the third amplifier is responsible for the high gain [1].

Initially let us assume $V_1=V_2$ (common mode) then I=0, $V_2'=V_2$ and $V_1'=V_1$. If $V_1 \neq V_2$ then $V'_2-V'_1 > V_2-V_1$.

The differential amplifier output is given as:

$$Vo = \frac{R2}{R1}V2' + \left[\left(1+\frac{R2}{R1}\right).\left(\frac{R2}{R1+R2}\right).V1'\right]$$

Taking R_2/R_1 common in the above equation. We have:

$$Vo = \frac{R2}{R1}\left[-V2' + \left[\left(\frac{R1}{R2}+1\right).\left(\frac{1}{1+\frac{R1}{R2}}\right).V1'\right]\right]$$

$$Vo = \frac{R2}{R1}(V1' - V2')$$

From the circuit we get,

$$I = \frac{V1 - V2}{R}$$

Also,

$$V1' = IR' + V1$$

And,

$$V2' = -IR' + V2$$

Substituting the value of I in V'_1 and V'_2 we get:

$$V1' = \left(\frac{V1 - V2}{R}\right).R' + V1$$

$$V2' = -\left(\frac{V1 - V2}{R}\right).R' + V2$$

Using the above in the output voltage equation.
We obtain:

$$Vo = \frac{R2}{R1}.\left[\frac{V1 - V2}{R}.R' + V1 + \frac{V1 - V2}{R}.R' - V2\right]$$

Simplifying, we get:

$$Vo = \frac{R2}{R1}.\left[2.\frac{V1 - V2}{R}.R' + (V1 - V2)\right]$$

$$Vo = \frac{R2}{R1}.\left(1 + \frac{2.R'}{R}\right).(V1 - V2)$$

III. DESIGN OF AN OPERATIONAL AMPLIFIER

This section describes the design of operational amplifier which is basic block in the instrumentation amplifier. The figure.4 represents the transistor level diagram of two stage operational amplifier.

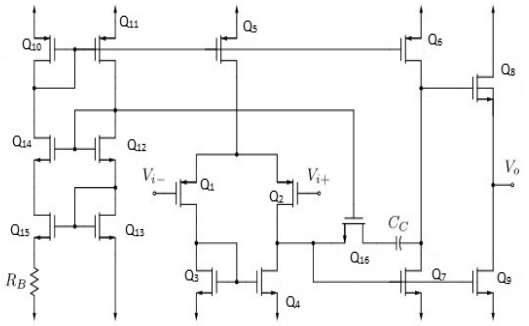


Figure.4: CMOS realization of a two-stage amplifier

A) Operational Amplifier Gain:

For low frequency applications, the overall gain is one of the most critical parameters. The gain of the first stage can be derived as follows:

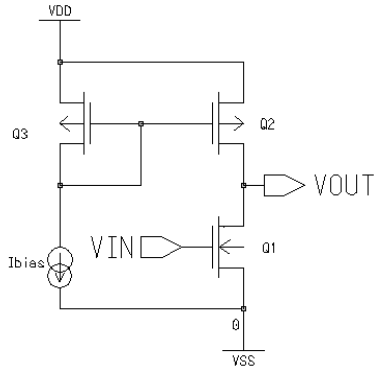


Figure.5: Common-source amplifier with a current mirror active load.

A small signal equivalent circuit for low frequency analysis for the common source amplifier is as follows:

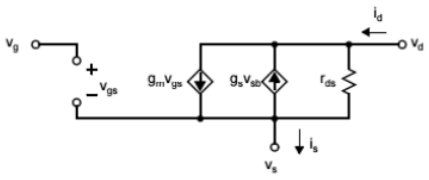


Figure.6: Small signal equivalent circuit for the common-source amplifier.

V_{in} and R_{in} are the Thevenin equivalent of the input source. It is assumed that the bias voltages are such that both the transistors are in active region. The output resistance, R_2, is made up of the parallel combination of the drain-to-source resistance of Q_1, that is, r_{ds1}, and the drain-to-source resistance of Q_2, that is r_{ds} [2].

Using small-signal analysis, we have $V_{gs1}=V_{in}$ an we have,

$$Av = \frac{Vout}{Vin} = -gm1.R2 = -r_{ds1} \| r_{ds2})$$

gm1 is given by,

$$gm = \frac{\partial Id}{\partial Vgs}$$

I_d in the active region is given by,

$$Id = \frac{\mu nCox}{2}\left(\frac{W}{L}\right)(Vgs - Vtn)$$

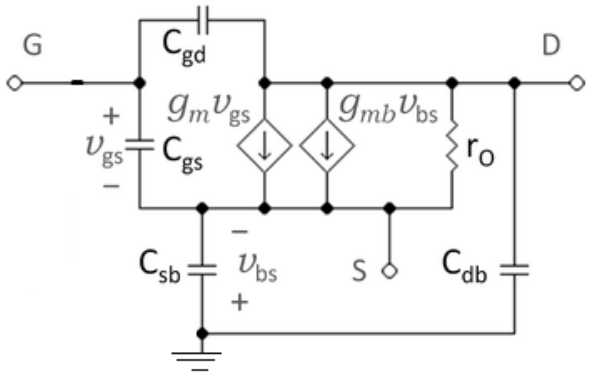


Figure 7: Small signal model for a MOS transistor in active region.

$$gm = \frac{\partial Id}{\partial Vgs} = \mu nCox\frac{W}{L}(Vgs - Vtn) = \mu nCox\frac{W}{L}Veff$$

Equivalently, we get:

$$gm = \mu nCox\frac{W}{L}Veff$$

Where, the effective gate-to-source voltage, Veff is defined as,

$$Veff = Vgs - Vtn$$

We now understand that the trans-conductance is directly proportional to Veff. It is desirable to express g_m in terms of I_D rather than Vgs, so we have:

$$Vgs = Vtn + \sqrt{\frac{2.Id}{\mu nCox(W/L)}}$$

From the above, we get:

$$Veff = Vgs - Vtn = \sqrt{\frac{2.Id}{\mu nCox(W/L)}}$$

Substituting Veff in gm and simplifying, we get:

$$gm = \sqrt{2.\mu n.Cox.\frac{W}{L}.Id}$$

Thus, the transistor trans-conductance is proportional to for a MOS transistor. Therefore, gm1 in gain of the first stage is given by,

$$gm1 = \sqrt{2.\mu n.Cox.\left(\frac{W}{L}\right).Id1} = \sqrt{2.\mu n.Cox.\left(\frac{W}{L}\right).\frac{Ibias}{2}}$$

Also, an approximation to the finite output impedance, r_{ds1}, of transistor Q1 is given by,

$$rds1 \approx \alpha\frac{Li}{Idi}\sqrt{Vdgi + Vti}$$

Where, α is a technology-dependent parameter of around 5 x 10^6.The second gain stage is simply a common-source gain stage with a p-channel active load Q6. Its gain is given as follows:

$$Av2 = -gm7(rds6||rds7)$$

The third stage is a common-drain buffer stage. This stage is often called a source follower, because the source voltage follows the gate voltage of Q8, except for a level shift. The gain of this source-follower is given by,

$$Av3 \approx \frac{gm8}{GL + gm8 + gds8 + gds9}$$

Where, G_L is the load conductance being driven by the buffer stage. When it is not possible to tie the substrate of Q8 to its source, as is the case when an n-well process is used, then the gain of the buffer stage is given by,

$$Av3 \approx \frac{gm8}{GL + gm8 + gs8 + gds8 + gds9}$$

Where, g_s is a body-effect conductance and is given by,

$$gs = \frac{gm.\gamma}{2.\sqrt{Vsb + 2\emptyset F}}$$

Voltage VSB is the source-to-substrate voltage and γ is the body-effect constant and 2□F is twice the difference between Fermi level in the bulk and the Fermi level of intrinsic silicon. Thus gs is around gm/5.

B) Frequency Response

The frequency response of the two-stage operational amplifier at frequencies where the compensation capacitor, C_C, has caused the magnitude of the gain to begin to decrease. There are some assumptions considered, first, all the capacitors are ignored except the compensation capacitor. Second, we assume that the transistor isn't

present. This transistor operates as a resistor, which is included to achieve lead compensation and it has an effect only around the unity-gain frequency of the operational amplifier [3].

The second stage introduces a capacitive load on the first stage due to the compensation capacitor, C_C. using Miller's theorem, the equivalent capacitance C_{eq} at node V_1 is given by:

$$Ceq = Cc(1 + A2) \approx Cc.A2$$

The gain in the first stage can be found using the small-signal model, and it results in:

$$A1 = \frac{V1}{Vin} = -gm1.Zout1$$

Where,

$$Zout1 = rds2||rds4||\frac{1}{s.Ceq}$$

For mid-band frequencies, the impedance of Ceq dominates, and the above equation can be written as follows:

$$Zout1 \cong \frac{1}{s.Ceq} \cong \frac{1}{s.Cc.A2}$$

For the overall gain, we have:

$$Av(s) \cong \frac{Vout}{Vin} = A3.A2.A1 \cong A3.A2.\frac{gm1}{s.Cc.A2}$$

If we assume A3=1, then the overall gain simplifies to:

$$Av(s) = \frac{gm1}{s.Cc}$$

Using this equation the approximation of unity-gain frequency can be done. To find unity-gain frequency, ωta, we set, |Av (jωta)|=1. Thus we obtain the following relation:

$$\omega ta = \frac{gm1}{Cc}$$

C) Slew Rate

Slew rate is one of the important high frequency parameter of an operational amplifier. The slew rate is the maximum rate at which the output changes when input signals are large. When the operational amplifier is limited by its slew rate because a large input signal is present, all of the bias current of Q5 goes into either Q1 or Q2, depending on whether Vin is negative or positive. When Vin is a large positive voltage, the bias current, ID5, goes entirely through Q1 and also goes into current mirror pair, Q3, Q4. Thus, the current coming out of the compensation capacitor, CC, is simply equal to ID5 since Q2 is off. When Vin is a large negative voltage, the current mirror pair, Q3 and Q4 is shut off because Q1 is off, and now the bias current, ID5, goes directly into CC. in either case, the maximum current entering or leaving CC is simply the total bias current[4], ID5.

Slew rate is the maximum rate that V2 can change. Recalling that Vout≈V2, we have:

$$Slew\ rate = \frac{dVout}{dt} = \frac{ICc|max}{Cc} = \frac{Id5}{Cc}$$

Where, we used the charge equation q=CV, which leads to I=dq/dt=C(dV/dt). Since ID5=ID1, we can rewrite slew rate as follows:

$$Slew\ rate = \frac{2.Id1}{Cc}$$

Where ID1 is the original bias current of Q1 with no signals present. Also, using the unity-gain frequency equation, we get CC=gm1/ωta, substituting this in the above equation[5]. We get,

$$Slew\ rate = \frac{2.Id1.\omega ta}{gm1}$$

Recalling the value of g_{m1},

$$gm1 = \sqrt{2.\mu n.Cox.\left(\frac{W}{L}\right).Id1}$$

Using this in the slew rate equation we have:

$$Slew\ rate = \frac{2.Id1}{\sqrt{2.\mu p.Cox.(W/L).Id1}}.\omega ta = Veff.\omega ta$$

Where,

$$Veff = \sqrt{\frac{2.Id1}{\mu nCox(W/L)}}$$

D) Systematic Offset Voltage

When designing the two-stage operational amplifier, it is possible that the design will have an inherent (or systematic) input-offset voltage. To ensure that no systematic input-offset voltage exists, when the differential input voltage is zero (i.e., when Vin+=Vin-), the output voltage of the first stage, VGS7 , should be that which is required to make ID7 equal to its bias current, ID6.

VGS7 is given by,

$$Vgs7 = \sqrt{\frac{2.Id6}{\mu n.Cox.(W/L)7}} + Vtn$$

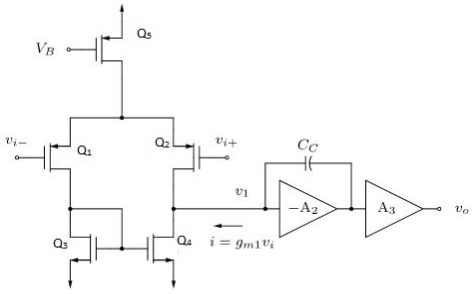


Figure.8: Input and gain stages of the two-stage op-amp.

When the differential input voltage is zero, the drain voltages of both Q3 and Q4 are equal by arguments of symmetry[6]. Therefore, the output voltage of the first stage, VGS7, is given by:

$$V_{GS7}=V_{DS3}=V_{GS4}$$

This value is the voltage necessary to cause ID7 to be equal to ID6. If this is not achieved, then the output of the second stage (with Q6, Q7) would clip at either the negative or positive rail since this stage has such a high gain[7]. However the gate-to-source voltage of Q4 is given as follows:

$$Vgs4 = \sqrt{\frac{2.Id4}{\mu n.Cox.(W/L)4}} + Vtn$$

Equating V_{GS7} and V_{GS4}, we get:

$$\sqrt{\frac{2.Id4}{\mu n.Cox.(W/L)4}} = \sqrt{\frac{2.Id6}{\mu n.Cox.(W/L)7}}$$

Simplifying, we have:

$$\frac{Id4}{(W/L)4} = \frac{Id6}{(W/L)7}$$

This equality, when the current density of Q_4 is equal to the current density of Q_7, guarantees that they both have the same effective gate-source voltages.
Since,

$$\frac{Id6}{Id4} = \frac{Id6}{Id5/2} = \frac{(W/L)6}{(W/L)5/2}$$

The necessary condition to ensure that no input-offset voltage is present it,

$$\frac{(W/L)7}{(W/L)4} = 2\frac{(W/L)6}{(W/L)5}$$

Using the above design equations, the width and length of the transistors have been found for the values assumed R_b=4.499KΩ and C_C=1.2pF.

Table .2: The W/L ratios of transistors in op-amp.

TRANSISTOR	WIDTH	LENGTH
Q_1	200µm	600nm
Q_2	200µm	600nm
Q_3	32µm	750nm
Q_4	32µm	750nm
Q_5	75µm	600nm
Q_6	312.52µm	600nm
Q_7	266.67µm	750nm
Q_{10}	11.66µm	1µm
Q_{11}	11.66µm	1µm
Q_{12}	3µm	1µm
Q_{13}	3µm	1µm
Q_{14}	3µm	1µm
Q_{15}	12µm	1µm
Q_{16}	18.905µm	700nm

IV. IMPLEMENTATION OF INSTRUMENTATION AMPLIFIER

An integrated chip shortly called as an IC can be designed in two ways. One is called Full-custom design and other the semi-custom design. The choice of particular design style depends on the factors which are required. Full costumed IC's consists of making the IC from the scratch. The basic building block of any chip is an interconnection of transistors and in full custom design each and every transistor is manually designed. This account for so much of time but advantage comes to be that circuit can be made for desired performance. Whereas in semi costumed IC's the design is mapped to pre fabricated gates whose technology is fixed [9].

The Figure.9 and Figure.10 represents the CMOS transistor level diagrams of two stage operational amplifier and instrumentation amplifier , respectively which are implemented in 180nm CMOS technology by CADENCE tools.

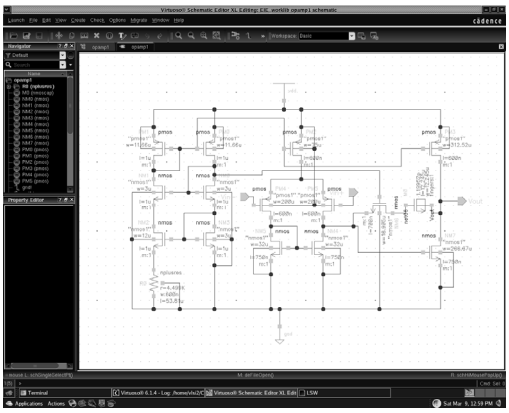


Figure.9: Schematic of operational amplifier

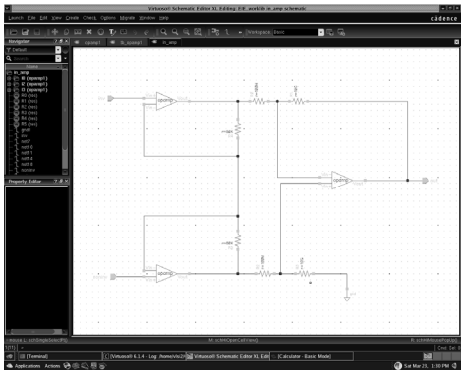


Figure.10: Schematic of instrumentation amplifier

V. SIMULATION RESULTS

In this section, the test bench set up for the instrumentation amplifier and simulation results are shown.

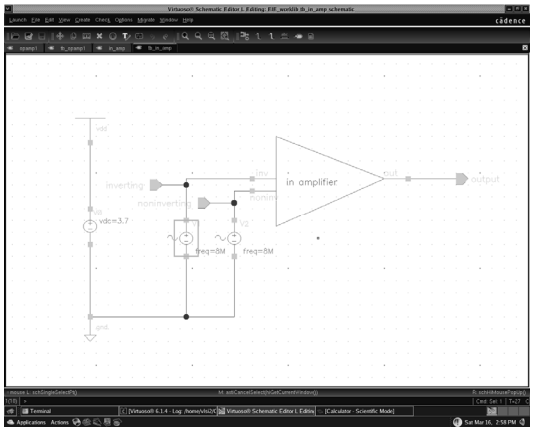


Figure.11: Test bench of instrumentation amplifier

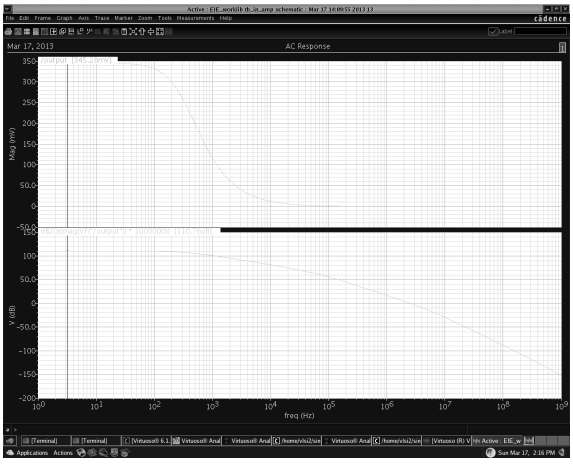


Figure.12: Simulation result of instrumentation amplifier

The above figure shows the gain of an instrumentation amplifier, which is around 111dB. The gain is obtained using AC analysis i.e. frequency response of the circuit. We also use the on screen calculator for the calculation of gain [8].The same instrumentation amplifier test bench is used to find the CMRR. A dc supply to this is 1.8V.

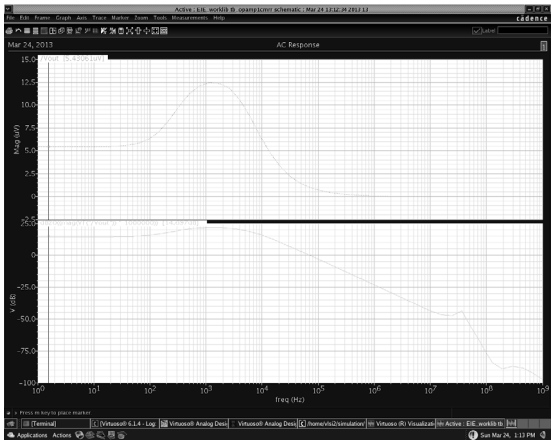


Figure13: CMRR response of an instrumentation amplifier.

The AC analysis is done for the circuit used earlier (test bench), the graph is obtained i.e. the frequency response and the CMRR is obtained by using the on screen calculator. The value of CMRR is around 15dB.

We know that offset is that voltage seen across the output terminals when no input is applied. So to find out the offset voltage, we give zero input and observe the output. The same has been implemented in a circuit form.

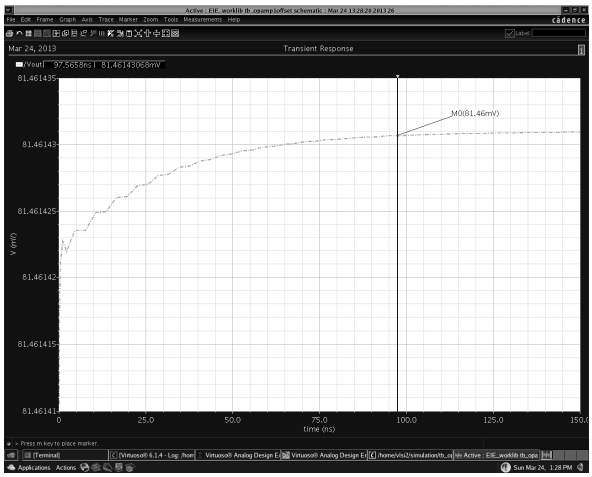


Figure.14: Offset voltage of an instrumentation amplifier

The response for the earlier circuit to find the offset voltage. We perform the transient analysis and find out the offset voltage. For the instrumentation amplifier it was found to be around 81.5mV, comparatively less than that of the operational amplifier.

To find PSRR we give both AC and dc supply to the entire test bench. One input of the instrumentation amplifier is supplied with dc voltage and the other is fed to the output pin.

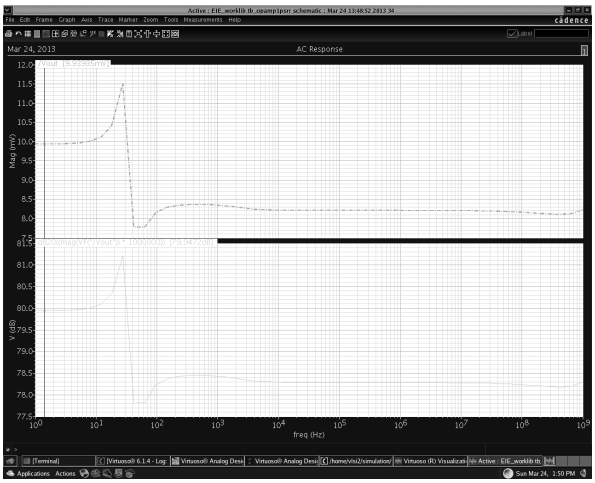


Figure.15: PSRR response of instrumentation amplifier

Using the PSRR test bench and performing the AC analysis i.e. finding the frequency response we obtain a plot. Using the onscreen calculator we can find out the ratio which was found to be around 80dB. The on screen calculator is used to find out the power dissipation and it is found to be around 5mW for the instrumentation amplifier.

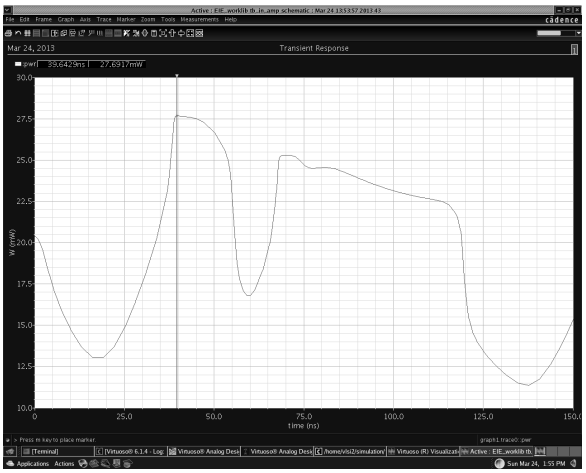


Figure.16: Power dissipation of an instrumentation amplifier

The following table represents the important parameters of Instrumentation amplifier and two stage operational amplifier, which are calculated from the simulation results.

Table.3: Calculated parameters

PARAMETER	OPERATIONAL AMPLIFIER	INSTRUMENTATION AMPLIFIER
Gain	82dB	111dB
CMRR	67dB	15dB
Offset	683.91mV	79.94mV
PSRR	79.99dB	79.94dB
Power dissipation	5.05mW	15.34mW

VI. CONCLUSIONS

The instrumentation amplifier successfully designed and achieving a gain of 110dB. A two stage operational amplifier has been used to implement the instrumentation amplifier. And the individual gain of the operational amplifier is 81dB. Since operational amplifier is the most basic amplifier used in signal conditioning circuits and this can be used for different applications. We have 180nm CMOS technology to implement it but, we can use further more advanced technologies like the 90nm or 45nm CMOS technology where the chip size is reduced thus power consumption is also reduced.

This instrumentation amplifier can be used for all the applications where the signals have very low amplitude. Usually, low amplitude signals are bio-signals and this instrumentation amplifier can be used to amplify bio-signals.

REFERENCES

[1] Roland D. Thomas and Albert J. Rosa, *The Analysis and Design of Linear Circuits*, United States of America, John Wiley and Sons, Inc., 2001.

[2] Pallas-Areny, Ramon and Webster, John G. Common Mode Rejection Ratio in Differential Amplifiers" *IEEE Transactions On Instrumentation and Measurement*,Vol. 40, No 4, August 1991, pp 669-676.

[3] D. Rod White "Phase Compensation of the Three Op Amp Instrumentation Amplifier" *IEEE Transactions on Instrumentation and Measurement*, Vol. IM-36, September 1987.

[4] Grasso, A.D., Palumbo, G., Pennisi, S., "Advances in Reversed Nested Miller Compensation," IEEE Transactions on Circuits and Systems-I, Regular Papers, vol.54, no.7, July 2007.

[5] Shen, Meng-Hung et al., "A 1.2V Fully Differential Amplifier with Buffered Reverse Nested Miller and Feedforward Compensation," IEEE Asian Solid-State Circuits Conference, 2006, p 171-174.

[6] Jungwon Han, Kwisung Yoo 'A low- power CMOS Limiting Amplifier Using Negative Impedance Compensation and It's Application' IEEE Transactions On VLSI Systems, Vol. 20, No. 3, March 2012 .

[7] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955.

[8] Aspaik, Liu 'A CMOS Instrumentation Amplifier With 90-dB CMRR at 2-MHz Using Capacitive Neutralization: Analysis, Design Considerations, and Implementation' IEEE Transaction on circuits and system, vol 50, No 4, Apr 2011.

[9] A. A. Khan, M.A. Al-Turaigi, and M. A. El-Ela, "An improved current-mode instrumentation amplifier with bandwidth independent of gain," IEEE Trans. Instrum. Meas., vol. 44, pp. 887–891, Aug. 1995.

Web Surfer Tracking using Big Data Technologies

B. B. Jayasingh,
Department of IT, CVR College of Engineering, Ibrahimpatan, Hyderabad, India
Email: bbjayasingh9@rediffmail.com

Abstract- **Log files are semi-structured on the web server-side. Computer generates huge flat text files as log. Therefore, Hadoop file systems are suitable to store as text files. The MapReduce techniques of the Hadoop works well in distributed clusters as it process one line at a time. NASA web server log is considered as input to system in order to perform the session identification task. A Statistical report is produced based on total count of visit per hour, per day, and per date. R language is used in order to find the user sessions and analyzed rigorously. In this paper, the distributed cluster is considered in Hadoop for the session identification in the web log. The analyst loads the log file and does analysis using MapReduce and generating statistical reports. The system developed to analyze the traffic from the log file based on various parameters, such as hours of a day, days of a week, days in a month etc. The system generates the detailed information about the location of a user that includes the country, state, city, zip code and also the co-ordinates of the location (latitude and longitude) by making use of the IP address field of a web server log.**

Index terms **– Web log, Web Mining, Big Data, Hadoop, Map Reduce, R Tool.**

I. INTRODUCTION

Web mining is the popular research area as the web generates a huge unstructured data lead to big data. The complexity of the web data requires effective search tools to find relevant information easily and precisely. The amount of information available in the web log is difficult to predict without a tool. How to predict the users' behavior is a matter of concern for the service provider and how to reduce the load on traffic as the personal information are more. The service providers must design the website which must accommodate various independent users.

The importance of user data is well known to companies like Facebook and Twitter though they have the biggest and fastest growing data repository in the world. Their size and diversity is required to be mined for useful information. The algorithms and the data analysis tools for inspecting the data are needed for decision making. The decision makers extract the useful information by considering specific digital footprints. The digital footprints are the representative of collective user behavior. The decision makers analyze a large amount of these traces to find common patterns, make better predictions, build smarter products, extract user models, and gain a better understanding of the dynamics of human behavior.

The data is surely growing in size, but also in complexity as it shows up in different formats and from different sources that are hard to integrate, and in dynamicity as it arrives continuously, changes rapidly and needs to be processed as fast as possible. Some vendors are using increased memory and powerful parallel processing to crunch large volumes of data extremely quickly. Another method is putting data in-memory but using a grid computing approach, where many machines are used to solve a problem. Both approaches allow organizations to explore huge data volumes and gain business insights in near-real time. On the other hand, it is a challenge [9]. Current methodologies are often not suitable to handle huge datasets, so new solutions are needed.

The World Wide Web is an interdisciplinary part of human life thereby the volume of data becomes larger and larger time to time. Big data [2] is when the size of the data itself becomes part of the problem and traditional techniques for working with data run out of steam. big data is data whose size forces us to look beyond the tried-and-true methods that are prevalent at that time. This means that we can call big an amount of data that forces us to use or create innovative methodologies. Visualization helps organizations perform analyses and make decisions much more rapidly, but the challenge is going through the sheer volumes of data and accessing the level of detail needed, all at a high speed.

In the business intelligence, Big data can handle petabytes or terabytes of data in a reasonable amount of time. Big data uses Hadoop framework for data intensive distributed applications. The main principle of hadoop is moving computations on the data rather the moving data for computation. Hadoop is used to breakdown the large number of input data into smaller chunks and each can be processed separately on different machines. To achieve parallel execution, Hadoop implements a MapReduce programming model. MapReduce a java based distributed programming model consists of two phases: a massively parallel "Map" phase, followed by an aggregating "Reduce" phase. MapReduce is a programming model and an associated implementation for processing and generating large data sets [10].

This paper uses the web logs of NASA website accessed by the user which is freely available, to identify the session. It also applies Hadoop framework in turn MapReduce technique to process the web log. The file size of the log is 550MB and the dataset are collected from various time period of the same year. The identified session is analyzed based on hour, day, date and number of times visited using R tool.

II. LITERATURE SURVEY

There are variety of Web logs in the world and The data has been so large that it becomes difficult to analyze it with the help of our traditional mining methods. Therefore the Big data term has been introduced that exceeds the processing capability [11]. Now, Data mining techniques are replaced with Big Data Mining to discover the usage patterns from the logged data. To understand customer behavior, evaluate the effectiveness of a particular website and the user, it is now big data mining techniques that plays an important role. It has three main key characteristics (1) volume (2) velocity and (3) variety. Therefore, Volume is the size of data which is now larger than terabytes and petabytes. So it is very difficult to analyse using conventional methods due to large scale. The Velocity is the pre-defined period of time to mine large amount of data using big data technology. The traditional methods of mining are not an appropriate solution as it may take huge time to mine such a volume of data. The Variety is the various heterogeneous sources the data comes from and forms a Bigdata. But the Bigdata is designed to handle structured, semi-structured as well as unstructured data which is not designed in the traditional methods that handle only structured data and not such large volume.

Hadoop [4] proposed the smart miner framework that extracts the user behaviors from web log. Web log contains data not only structured traditional relational data, but also semi-structured and unstructured data come from a variety of sources and in a variety of types. There are more focused research towards web log mining using big data technologies. Such a framework [5] is used the smart session construction to trace the frequent user access paths.

A traditional data warehouse cannot accommodate the data generated by machine such as click stream logs, email logs. Though the data are of unstructured that are larger in volume in comparison with human generated data. So Big data is the only alternative for the recent trend to store and analyze. The work embodied in this paper [6] is a generic log analyzer framework for different kinds of log file. The other work proposes the train model where the data is stored in HDFS and the test model categories the text document is the Parallelization of Genetic

Algorithm (PGA) is suggested [7] that uses OlexGA package for classifying the document.

There are two kinds of node in a cluster where one node acts as both master and slave and the other node acts as a slave and the data is transferred in 100Mb/s speed. There are five daemons such as namenode, datanode, jobtracker, tasktracker and Secondary namenode that contains a master node. A slave node contains tasktracker and datanode daemons and the master node contains the IP address of the slave. The slave node is identified by the master using its ip address. The master node read the log files that are of block size of 64MB, so it is default to save in hard drive. A framework for unstructured data analysis was proposed [8] by using big data of public tweets from twitter. The tweets are stored in Hbase using Hadoop cluster through Rest Calls and text mining algorithms that are processed for data analysis.

III. HADOOP FRAMEWORK

Hadoop framework [11], In pseudo distributed mode contains all the five daemons run on local machine simulating a cluster. It process mostly unstructured text files, so the text files are generated and stored in the HDFS after applying data cleaning step. The cleaned web log data is used to analyze the session identification, unique user and unique URLs. HDFS stores large files across multiple machines typically in the range of gigabytes to terabytes. It achieves reliability by replicating the data across multiple hosts. Hadoop implements a computational paradigm known as MapReduce.

MapReduce [10] is a computational paradigm designed to process very large sets of data in a distributed fashion. The model is based on the concept of breaking the data processing task into two smaller tasks of mapping and reduction. During the map process, a key-value pair in one domain is mapped to a key-value pair in another pair, where the 'value' can be a singleora list of multiple values. The keys from the mapping process are then aggregated and the values for the same key combined together. This aggregated data is then fed to the reducer (one call per key) and the reducer then processes this data to produce a final value. The list of all final values for all the keys is the result set.

The key issue in breaking a problem into the MapReduce model is that the map and reduce operations can be performed in parallel on different keys, without the results of one operation affecting the other. This independence of results allows the map/reduce tasks to be distributed in parallel to multiple nodes, which can then perform the respective operations independent of each other. The final results are then aggregated together to produce the final result list.

IV. WEB LOG ANALYSIS

User logs are collected by the web server and typically include IP address, page reference and access time. Mining web data provides a lot of information, which can be better understood with visualization tools. This makes concepts clearer than is possible with pure textual representation. Hence, there is a need to develop tools that provide a graphical interface that aids in visualizing results of web mining. Some of the most prominent technologies are [3] NoSQL database.

Finding hidden patterns in the large database needs analytical techniques. There are so many software tools available for predictive analysis including big data analytics to find useful information. The recent technologies focus on big data to analyze the logged data to track the user's behavior. We try to track the user behavior by analyzing the logged datasets that are in semi structured fashion maintained by Hadoop. We took a log

record as an example and analyzed all the fields incorporated in the table given herewith.

An example of log record is 125.125.125.125 - uche [20/Jul/2008:12:30:45 +0700] "GET /index.html HTTP/1.1" 200 2345

Field name	Example value	Description
host	125.125.125.125	IP address or host name of the HTTP client that made the request
identd	-	Authentication Server Protocol (RFC 931) identifier for the client; this field is rarely used. If unused it's given as "-".
username	uche	HTTP authenticated user name (via 401 response handshake); this is the login and password dialog you see on some sites, as opposed to a login form embedded in a Web page, where your ID information is stored in a server-side session. If unused (for example, when the request is for an unrestricted resource) it's given as "-".
date/time	20/Jul/2008:12:30:45 +0700	Date then time then timezone, in the format [dd/MMM/yyyy:hh:mm:ss +-hhmm]
request line	GET /index.html HTTP/1.1"	The leading line of the HTTP request, which includes the method ("GET"), the requested resource, and the HTTP protocol version
Status code	200	Numeric code used in the response to indicating the disposition of the request, for example to indicate success, failure, redirect, or authentication requirement
bytes		Number of bytes transferred in the body of the response

V. OUTPUT VISUALIZATION

The system can be used to visualize the traffic from a log file based on various parameters, such as hours of a day, days of a week, days in a month etc. The system generates the detailed information about the location of a user, who requests the web site for a URL. The information includes the country, state, city , zip code and also the co-ordinates of the location(latitude and longitude) by making use of the IP address field of a web server log. The system provides the user with the option of visualizing the results of log analysis. The system plots the geolocation of all the ip addresses present in the log file on a world map using R programming. This can be used to analyze the traffic distribution based on the location of the user. The system generates line graphs to visualize the traffic distribution for hours of the day, days of a week and days of a month. These graphs can be used to detect patterns in traffic distribution and those patterns can be applied for business purposes.

As more and more businesses are discovering, data visualization is becoming an increasingly important component of analytics in the age of big data. Plotting points on a graph for analysis becomes difficult when dealing with extremely large amounts of information or a variety of categories of information. So we use a free statistical computing tool and plot the graphs that are through R language.

The following R script in Fig. 1 takes the hours and requests as input and generates a line plot with hours on x-axis and requests on y-axis.

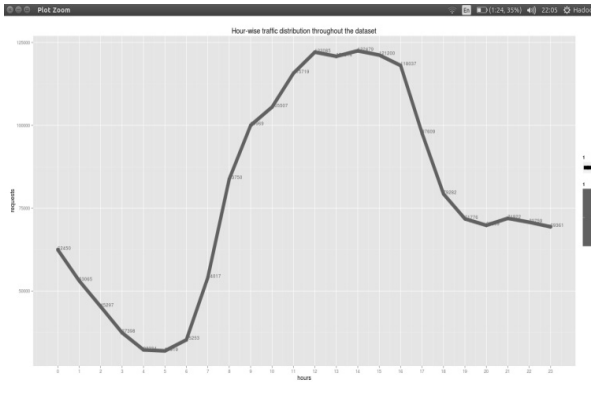


Fig. 1 Traffic distribution for hours of the day

The following R script in Fig. 2 takes the dates and requests as input and generates a line plot with dates on x-axis and requests on y-axis

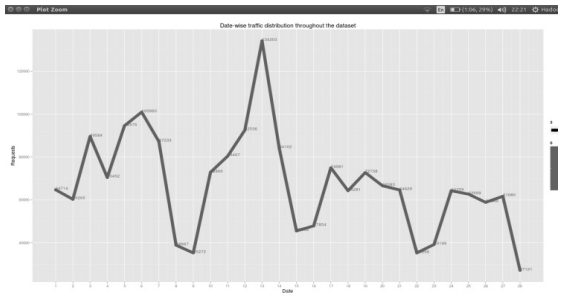


Fig. 2 Traffic distribution for Dates of the month

The following R script in Fig. 3 takes the days and requests as input and generates a line plot with days on x-axis and requests on y-axis

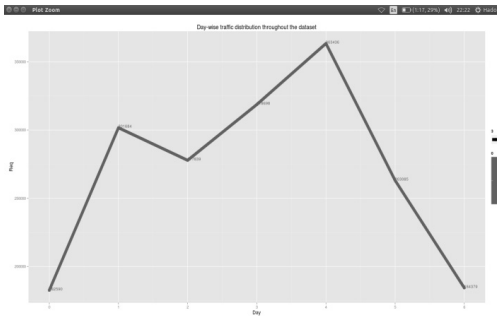


Fig. 3 Traffic distribution for Days of the week

The following R script in Fig. 4 takes the co-ordinates of the geolocation of the IP addresses from which the requests have been generated. The co-ordinates are plotted on a world map, visualising the distribution of traffic to the web-site.

Fig. 4 Traffic distribution for location of the user

VI. CONCLUSIONS

The importance of the Big data technology and Hadoop framework is discussed in this paper. The Map Reduce techniques are implemented and tracked for the web surfer information. Our algorithms process the NASA web server logs using MapReduce task to produce a statistical report based on total count of visit per hour, per day, and per date. R programming language tool is used to visualize the details about the web surfer.

REFERENCES

[1] "G. C Bodyan, T.V Shestakov, "Web Mining in Technology Management, Engineering Universe for Scientific Research and Management, Vol 1 Issue 2, April 2009.

[2] "R. Verma, S. R. Mani, "Use of Big Data Tehnologies in Capital Markets," 2012 Infosys Limited, Bangalore, India
.

[3] "J. Manyika, B. Brown et.al, "Big Data: The next frontier for Innovation, Competition, and Productivity," McKinsey Global Institute, June 2011.

[4] "M. Ali , I. Hakki Toroslu, "Smart Miner: A New Framework for mining Large Scale Web Usage Data," WWW 2009, April 20-24. 2009 Madrid, Spain. ACM 978-1-60558-487-4/09/04.

[5] "S. Narkhede and Tripti Baraskar, "HMR Log Analyzer: Analyze Web Application Logs Over Hadoop MapReduce," International Journal of UbiComp (IJU) vol.4, No.3, July 2013.

[6] "M. Bhandare, Vikas Nagare et al., "Generic Log Analyzer Using Hadoop Mapreduce Framework," International Journal of Emerging Technology and Advanced Engineering (IJETAE), vol.3, issue 9, September 2013.

[7] "K Sharadchandra Rahate et al., "A Novel Technique for Parallelization of Genetic Algorithm using Hadoop," International Journal of Engineering Trends and Technology (IJETT), vol.4, issue 8, August 2013.

[8] "T. K. Das et al., "BIG Data Analytics: A Framework for Unstructured Data Analysis," International Journal of Engineering and Technology (IJET) vol 5, No 1, Feb-Mar 2013.

[9] "J. McKendrick, "Big Data, Big Challeneges, Big Opportunities: 2012 IOUG Big Data strategies survey," September 2012.

[10] "J. Dean and S. Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," Google, Inc

[11] "T. White, "Hadoop: The definitive Guide," Third Edition, ISBN: 978-1-449-31152-0-1327616795.

Virtualization Layer Security in Cloud Architecture

S. Jyothsna
Department of IT, CVR College of Engineering, Ibrahimpatnam, Hyderabad, India
Email:jyothsna.sundaragiri@gmail.com

Abstract: **Cloud Computing is a type of service based computing or utility computing. Cloud computing is based on virtualization technology. Virtualization architecture is categorized as hypervisor architecture, para virtualization and host based virtualization depending on the position of the virtualization layer. The hypervisor supports hardware-level virtualization on devices like CPU, disk, memory etc. The hypervisor provides hyper calls for the guest operating systems and applications. The architecture of processor supports the ability to run number of virtual machine instructions on one CPU after virtualization. Virtual machine instructions can be categorized as privileged and unprivileged instructions. Privileged instructions must be run in supervisor mode in Hypervisor. To provide reliable computing in cloud, virtualization layer security is a major concern. Hypervisor is a program responsible for allocation and de allocation of resources to each virtual machine (VM) connected to the cloud. Hypervisor security compromise affects all the privileged and sensitive instructions. Hypervisor is very small program compare to the operating system so it is easy to attack . A set of additional instructions must be added to control the hypervisor attacks and regular CPU state checking must be done in each virtual machine.**

Index terms: **cloud computing, virtualization, hypervisor, virtual machine, virtual machine monitor, hypervisor security**

I. INTRODUCTION

1.1 Cloud Computing

A Cloud can be considered as an enormous collection of resources. Cloud computing is providing easy accessibility for required services and users can also deploy applications at competetive costs. Large data centers provide services through virtualized cloud platforms. Cloud computing is providing services at infrastructure level are called as infrastructure as a service.

The cloud service provider (CSP) provides on demand provisioning of Hardware like processing power, I/O, large amounts of storage etc... By utilizing virtualization, each user accesses the services of cloud through a virtual machine, where number of virtual machines shares a single physical server. cloud computing provides a service oriented platform for cloud users.

1.2 Security in Virtualized Environment

The main idea of virtualization technology is to separate the hardware from software to improve efficiency of the system. The Cloud Service Provider provides the infrastructure usually in the form of virtual machines to manage compute resources efficiently. The virtualization software creates the images of computer system at application level is called a virtual machine(VM).Each user's application runs virtually on their VM, but VM cannot run instructions directly, most of the unprivileged VM instructions are executed directly on the host processor,these instructions must be handled carefully to avoid vulnerabilities and maintain stability of the system.
Virtualization Layer is the critical element of cloud computing, VMM (virtual machine monitor) is responsible for creation of virtual machines, resource allocation to virtual machines, isolation of virtual machines from each other etc.., virtualization improves resource utilization but security risk also increases, to provide an efficient cloud environment more standard security mechanism is required at virtualization layer.

1.2.1 Customers in Cloud Attack Vectors

The fig. 1 shows the customer environments 1 and 2 connected to the common cloud, where customer 2 being the attacker (red), and customer 1 being the victim (green).

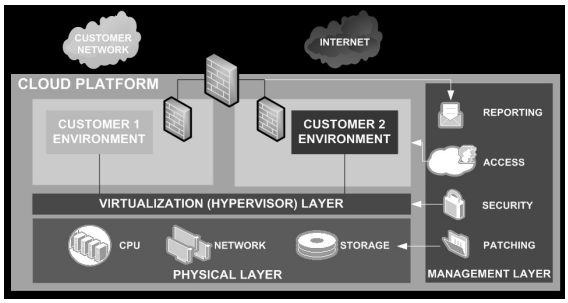


Fig. 1. Customer Environments Connected to Common Cloud

In this case, customer 1 also connects its own internal network (a situation common for many IaaS solutions). The parts shown in blue are managed by the Cloud Service Provider (CSP).
The attacks in cloud can be any of the following

1. Hypervisor/Virtualization Layer Attack, which is the focus of this paper.
2. Management layer vulnerabilities (blue)
3. External attacks from Internet (red)
4. Internal customer network threats

In multi processing environment more standard protection mechanism is mandatory to avoid system crash because direct accessing of hardware may take place by processes. Therefore, all processors have at least two modes, user mode and supervisor mode to ensure controlled access of critical hardware. There are relatively more layers in the machine stack in a virtualized environment so it is more difficult to make operating systems and applications run correctly.

II. HYPERVISOR ARCHITECTURE

2.1 Hypervisor

The hypervisor is a virtual layer between the physical hardware and its operating system, in cloud architecture virtualization software (ex. xen) is used to create virtual machines. Virtual machines may run on different operating systems such as linux and windows, i.e. different operating systems run on the same physical hardware simultaneously. converting portions of the real hardware into virtual hardware is the major responsibility of virtualization layer . Hypervisor may be placed above hardware as a separate layer or as part of real operating system.

The Hypervisor supports hardware-level virtualization on CPU, memory, disk and network interfaces. The hypervisor provides hyper calls for the guest operating systems shown in figure 2.

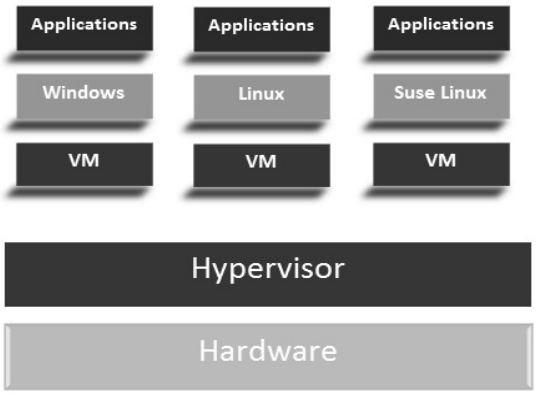


Fig. 2 Hypervisor Architecture

2.2 Hypervisor Design Goals

This virtualization layer is referred to as either the VMM or the hypervisor. Depending on the functionality, a hypervisor can assume micro- kernel architecture like the Microsoft Hyper-V or it can assume monolithic hypervisor architecture like the VMware for server virtualization.

Hypervisor provides

- Physical memory management
- Processor scheduling
- Converts physical devices into virtual resources

Virtual layer is responsible to distribute and assign resources to virtual machines in cloud environment. At the cloud layer, sets of virtual devices are sandboxed and offered to clients as if they were completely separated data centers. Thus, cloud computing is very dependent on the quality of the hypervisor security. The hypervisor provides a very strong barrier between virtual machines and thus between different customers.

2.3 Virtualization Networking

Xen Server and VMware ESX Server are virtualization platforms and supports a bridging mode which allows all domains to appear on the network as individual hosts. By using this mode, VMs can communicate with one another freely through the virtual network interface card and configure the network automatically.

In a network built with mixed nodes of host and guest systems, the normal method of operation is to run everything on the physical machine. When a VM fails, its role could be replaced by another VM on a different node, as long as they both run with the same guest OS. The Potential drawback is that a VM must stop playing its role if its residing host node fails.

VMs can be live-migrated from one physical machine to another. Virtual clusters can be applied in Cloud Platforms. Virtual clustering plays a key role in cloud computing. The virtual clustering provides dynamic resources that can be quickly put together upon user demand or after a node failure. When a VM runs a live service, the following three metrics must be considered.

1. Negligible downtime
2. Lowest network bandwidth consumption
3. Reasonable total migration time

Furthermore the migration must ensure that it should not disrupt other active services residing in the same host through resource contention.

A migrating VM should maintain all open network connections without relying on forwarding mechanisms on the original host or on support from mobility or redirection mechanisms. To enable remote systems to locate and communicate with a VM, each VM must be assigned a virtual IP address known to other entities. This IP address can be distinct from the IP address of the host machine where the VM is currently located. Each VM can also have its own distinct virtual MAC address. The VMM maintains a mapping of the virtual IP and MAC addresses to their corresponding VMs.

Xen as a VMM allows multiple operating systems to share x86 hardware in a safe and orderly fashion. Xen supports live migration. It is a useful feature and natural extension to virtualization platforms that allows for the transfer of a VM from one physical machine to another with little downtime of the services hosted by VM. Xen Hypervisor uses a send/recv model to transfer states across Virtual machines.

2.4. Preventing Hypervisor Attacks

Guest hopping and hijacking or VM root kits are the major attacks in virtual layer (hypervisor) along with buffer overflows, distributed denial of service attacks in a cloud environment, another type of attack is the man-in-the-middle attack for VM migrations. Virtualization enhances resource provisioning in cloud but virtual machines add an additional layer of software that could become a single point of failure.

To prevent from failures

- Use redundant utilities at multiple sites
- Alternate network connections
- Multiple databases at separate sites
- Data watermarking and user authentication...
- Trust delegation and negotiation

- All datacenters can be secured from distributed denial of service attack by distributed defence and internet worm containment.
- Fine grained access control at file or object level.
- Use double authentication, biometric identification, intrusion detection and disaster recovery etc... for privacy protection

III. IMPLEMENTATION OF HYPERVISOR

The Hypervisor is a program executed by the server. When hypervisor is executed it loads the client operating systems of the virtual machines. The hypervisor allocates the correct CPU resources, memory, bandwidth and disk storage space for each virtual machine.

The main function of the VMM is to virtualize the physical hardware of a host machine into virtual resources to be used by the virtual machines. This can be implemented at various levels.

There are two types of hypervisors:

1. Bare metal or native hypervisors
2. Embedded or hosted hypervisors

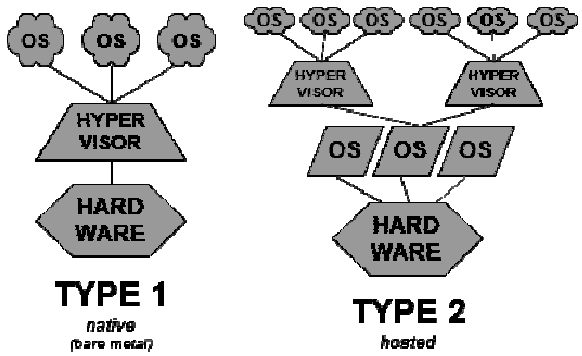


Fig. 3 Hypervisor Types

A virtual machine can create requests to the hypervisor through a variety of methods, including API calls.

3.1 Bare Metal or Native Hypervisors

This is also called as full virtualization. In this, non critical instructions are discovered and replaced with traps in to the VMM to be emulated by software. Non critical instructions do not control hardware or threaten the security of the system, but critical instructions do. Therefore running non-critical instructions on hardware improve efficiency by ensuring security.

The method used in this emulation is called binary translation. Binary translation is time consuming and increases the cost of memory usage. (To store translated instructions in cache). The Performance of full virtualization may not be ideal.

3.2 Host-Based Virtualization

In Host-Based virtualization, hypervisor is installed on top of host operating system. This host operating system is responsible for managing hardware. The guest operating systems are installed and run on top of the hypervisor. Dedicated applications may run on the virtual machines. Some other applications can also run with the host operating system directly.

Host based virtualization have following advantages

- The user can install VM architecture without modifying host operating system.
- Host based approach appeals to many host machine configurations.

The host based architecture has flexibility but the performance of the host based architecture also may be low because it involves four layers of mapping when an application requests hardware access.

IV. SECURITY AT HYPERVISOR LAYER

Many guest Operating systems can run on top of the hypervisor, one among those guest operating systems controls the others. This is called as Domain O, and the others are called Domain U in Xen. Domain O is the privileged guest operating system, which access the hardware resources directly, if Domain O is compromised, the hacker can control the entire system.

The processor power of the hardware is distributed among virtual machines belonging to different customers by the hypervisor layer. Hypervisor does not function natively on the host and can only access host resources through a separate control layer commonly named the virtual machine monitor (VMM).

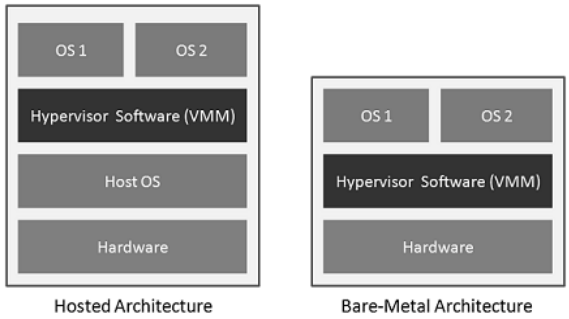


Fig. 4 VMM Architecture

There are four major levels in security management of hypervisor as mentioned below:

*Attribute based Authentication***:** Just-in-time tokens must be generated for users accessing applications of each guest OS separately for each access based on the attributes of application

Controlled Authorization: A model that includes pre-defined application roles and user roles and API based on user-role bindings only can be used to access.

Minimized Hypervisor Intervention: Hypervisor should be responsible for the creation of virtual machine, allocation of resources to virtual machine and shutting down of virtual machines only but runtime intervention of the hypervisor must be restricted i.e. resource allocation to virtual machines must be static so that security compromise of virtual machine monitor may not effect virtual machines.

Secure Code Execution: onion routing can be used to execute each virtual machine's code .Onion routing is a technique where messages are encapsulated in layers of encryption, The encrypted data is transmitted through a series of nodes from each virtual machine to datacenter in cloud.

Security Recommendations for Hypervisor Security

1. Resource allocation, processor core assignment and input/output calls must be done statically for each virtual machine to avoid active interaction with virtual layer.
2. Standard public key encryption techniques like RSA should be used to access management layer.
3. Every virtual machine data transmission activity must be controlled by VM Security Monitor (VSEM).
4. Clear segregation of security zones should be provided in the cloud environment for each virtual machine.
5. Standard encryption systems which can efficiently encrypt large volumes of data at boot level is preferred.
6. Memory regions of hypervisor can only be modified by instructions that are intended part of the hypervisor.

V. CONCLUSIONS

Cloud Computing is based on the virtualization technology, Virtualization provides more resources than actually available by multiplexing virtual machines into single hardware. Hypervisor based virtualization creates a new layer between the hardware and host operating system. The virtualization layer called virtual Machine Monitor (VMM) actually accesses the real hardware to provide infrastructure services to cloud users. Each client connected to the cloud, uses their own guest operating system for its functionality. VMM security compromise may affect all virtual machines. So that standard security mechanism must be provided to VMM to achieve reliable services of cloud.

REFERENCES

[1] Kai Hwang. Geoffrey C.Fox. Jack J.Dongarra (2012) - Distributed and cloud computing, Mogan Kaufmann publishers

[2] "Cloud_Security_demystified", CSC

[3] http://cybersecurity.mit.edu/2013/10/virtualization vulnerabilities-related-to-hypervisors/

[4] "Cloud Security Is Not (Just) Virtualization Security",Mihai Christodorescu, Reiner Sailer, Douglas Lee Schales.

[5] "Secure Virtualization for Cloud Environment Using Hypervisor-based Technology" - Farzad Sabahi, Member, IEEE

[6] "An Architecture for Concurrent Execution of Secure Environments in Clouds", Ramya Jayaram Masti, Claudio Marforio, Srdjan Capkun

[7] http://www.cpd.iit.edu/netsecure 08/ROBERT_RANDELL.pdf

[8] https://www.ernw.de/download/ERNW_DCVIH ypervisors To Clouds.pdf

Semantic Similarity Measurement between Words using Lexical Patterns

D. Hema Latha[1], D. Linga Reddy[2]
[1]Dept of Computer Science, Osmania University College For Women (OUCW)
Koti, Hyderabad, India
E mail : hlatha059@gmail.com
[2] Dept. of Physics, UCS, Osmania University, Hyderabad, India.
Email : dlreddy_phy@yahoo.com

Abstract - **Semantic similarity measurement between words is a tedious task in web mining, information extraction and natural language processing. The semantic similarity measurement between entities is required in Web mining applications such as community extraction, identification of relations etc. In this paper, the authors proposed an automatic approach to evaluate the logical or semantic similarity between words or entities with the help of web search engines. To describe distinct word co-occurrence measures and to integrate these with lexical patterns, page counts are used. In order to identify meaningful relationships between two given words, the authors proposed a new pattern extraction algorithm and a pattern clustering algorithm. Vector Support Machine (VSM) is used to acquire the optimal combination of page counts-based co-occurrence measures and lexical pattern clusters. The proposed method overcomes various previously proposed web-based similarity measures on the benchmark data sets that showed a high correlation with human ratings.**

Index terms - **Lexical Pattern, Web mining, Information Extraction**

I. INTRODUCTION

Semantic similarity, logical or meaningful association is a conception where a set of documents or terms within term lists are assigned a metric that is based on the likeness of their meaning or semantic content.

Given a group of words, similarity between two words can be calculated by the length of the shortest path that connects the two words in the group. Multiple paths may exist between the two words, if the word has many meanings. In such cases, only the shortest path between any two perceptions of the words is considered for similarity calculation.

The problem with this approach is that it depends on the concept that all the paths or links in the group represent a uniform distance. Resnik proposed a similarity calculation approach based on the content of the information. He described the similarity between two conceptions 'Y1' and 'Y2' in the group of words as the maximum of the information content of all concepts 'Y' that include both 'Y1' and 'Y2'. Then, the affinity between two words is described as the maximum of the similarity between any concepts that the words belong to. He used Word Net as the group or taxonomy and calculated the information content using the Brown corpus.

Li et al. combined structural semantic information from a lexical taxonomy and information content from a corpus in a nonlinear model. They proposed a similarity measure that uses short sighted length profundity depth and local compactness or density in taxonomy. Their experimental study reported a high Persuasion correlation coefficient of 0.8914 on the Miller and Charles example and reference or benchmark data set. They did not appraise or calculate their methodology in terms of similarities among named entities. Cilibrasi and Vitanyi proposed a distance metric between words using only page counts retrieved from a web search engine.

II. DESIRABLE FEATURES FOR RELATEDNESS MEASURE

Desirable features to measure semantic similarity in current Semantic Web applications.

1. Domain Independence: Presently, an increasing amount of online ontological and semantic data is available on the World Wide Web, enabling a new generation of semantic applications. If that kind of domain independent applications to be developed, this increasing heterogeneity should be dealt, without establishing the ontologisms to be accessed in advance.

2. Universality: The semantic measures, in the dynamic context of the Web, must be flexible, compatible and general enough to be used independently for their final purpose, and without relying on specific lexical resources or knowledge representation languages.

3. Maximum coverage: Maximum coverage of possible interpretations of the words must be warranted, in the context of web applications with no predefined domain. If it is limited to a particular knowledge source, such as WordNet2, or a certain set of ontology, then one is compelled to use those applications only.

III. LITERATURE SURVEY

Given a set of words, a direct approach to calculate similarity between two words is to compute the length of the shortest path connecting the two words in the set. If a word has many meanings (polysemy), then numerous paths might exist between the two words. In such cases, only the nearest path between any two perceptions of the words is considered for calculating similarity. A problem with this approach is that all paths or links in the group of words

represent a uniform distance. Resnik[8] proposed a logical resemblance measure using information content. He described the similarity between two conceptions 'Y1' and 'Y2' in the taxonomy or group as the maximum of the information content of all concepts 'Y' that include both 'Y1' and 'Y2'. Then, the affinity or closeness between two words is defined as the maximum of the similarity between any concepts that the words belong to. He used Word Net as the taxonomy; information content is calculated using the Brown corpus. Li et al. [9] combined organized meaningful information from a lexical taxonomy and information content from a corpus in a non-linear model. They proposed a similarity metric that uses short sighted length, profundity or depth, and local compactness or density in group. Their experimental study reported a high persuasion correlation coefficient of 0.8914 on the Miller and Charles [10] example, reference or benchmark data set. They did not appraise or calculate their methodology in terms of similarities among named entities. Lin [11] defined the similarity between two concepts as the information that is in common to both concepts and the information contained in each individual concept. Cilibrasi and Vitanyi [12] proposed a distance metric between words using only page counts retrieved from a web search engine.

IV HISTORY

The searching process shown in fig.1 gives the results based on the text search algorithms. Present all Search Engines on the web are based on the text search algorithms.

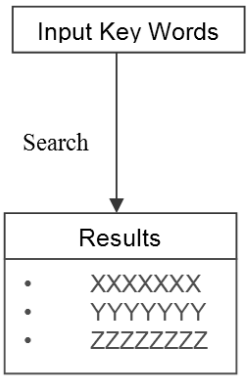


Figure 1. Generic search process

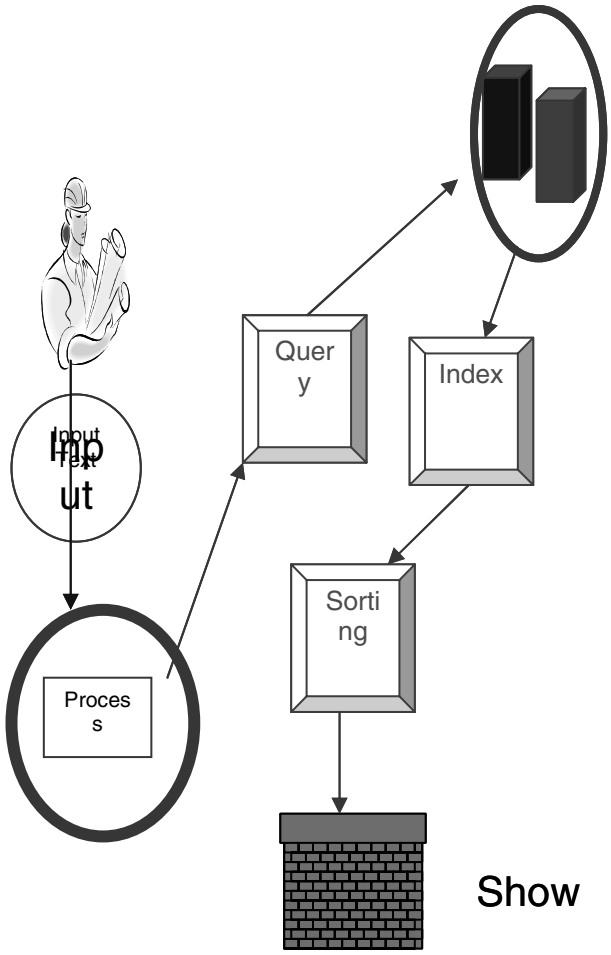


Figure 2. Generic Search Process Model

Fig. 1 shows the generic search process model

- A similarity measure can represent the similarity between two documents, two queries, or one document and one query
- It is possible to arrange the extracted documents in the order of presumed importance that is ranking the extracted documents in the order of presumed importance
- A similarity measurement is a strategy which computes the degree of similarity between a pair of text objects
- Many number of similarity measures are proposed in the literature, because the best similarity measure doesn't exist (yet!)

V. VECTOR-SPACE MODEL-VSM

1960s Salton provided Vector Space Model, which has been victoriously or favorably applied on SMART (a text searching system).

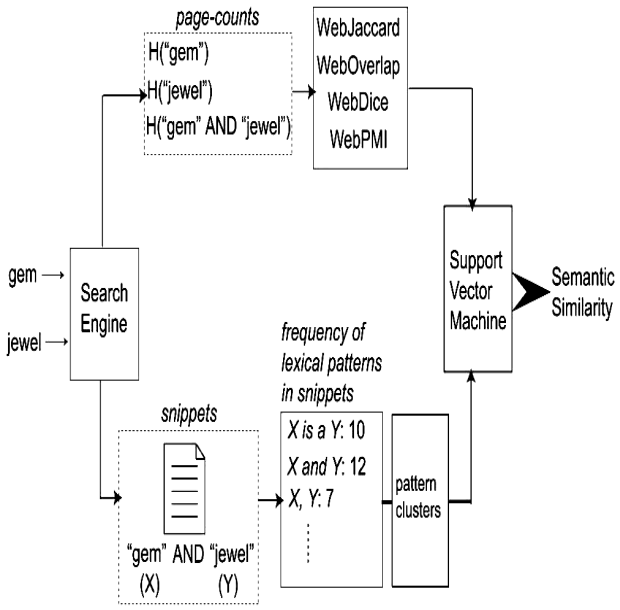


Figure. 3 Architecture of Vector Space Model

The Vector Space Model architecture is shown in fig.3

The vector space model procedure is categorized in to **three phases**.

- The **first** is the **document indexing** phase in which content bearing terms are extracted from the document text.
- An **indexed term weighting** is the second phase enhances the retrieval of document relevant to the user.
- The last phase provides ranking **to the document** with respect to the query according to a similarity measure.

From fig.3 'gem' and 'jewel' are given as input to web search engine. The search engine search for pages gem, jewel and gem as jewel in the form of page counts.
It displays text snippets 'gem' (X) and 'jewel' (Y).
The frequency of occurrence of lexical patterns is measured based on the snippets.
These patterns are clustered and fed to Vector Support Machine (VSM) as input.
The Vector Support Machine (VSM) allots weights and ranking is done based on the weights.

A. Document Indexing

In the document non significant words may appear, by using document indexing these non-significant words (function words) is removed so that the document is represented by content bearing words. This document indexing is done based on the frequent occurrence of the terms, where low frequency terms within a document are considered to be function words. Stopping list is used to remove high frequency words (stop words) which hold common words, which makes the indexing method language dependent. With the help of stop list 40% - 50% of the total number of words in a document is removed.

Probability Indexing is used which shows the statistical difference in the distribution of content bearing words, and the function words. Probabilistic indexing ranks the terms with respect to the term frequency in the entire collection. The function words are prototyped by a Poisson distribution in the overall documents, as content bearing terms cannot be prototyped. Recently, an automatic indexing method which uses serial clustering of words in text has been introduced. The value of such clustering is an indicator if the word is content bearing.

B.Term Weighting

Term weighting has been described in terms of recall and precision. There are three main components for calculating term weighting - term frequency component, collection frequency component and length normalization component. These three components are multiplied together to make the resulting term weight.

A common weighting scheme for terms within a document utilizes the frequency of occurrence as mentioned by Luhn. The term frequency for documents is generally used as the basis of a weighted document vector. It is also possible to use binary document vector, but the results are not that good when compared to the term frequency when using the vector space model.

Different weighting schemes are available to discriminate one document from the other. In general this component is called accumulation or collection frequency document. Most of them, e.g. the inverse document frequency, assume that the importance of a term is proportional with the number of document the term appears in. Experimentally it has been shown that these document discrimination factors lead to a more effective extraction, i.e., an improvement in precision and recall.

The third possible weighting factor is document length normalization factor. Lengthy documents have usually a much greater term set than small documents, which makes lengthy documents to be retrieved faster than small documents.

Experiments have been done on various weight schemes and achieved best results, with respect to recall and precision, are acquired by using term frequency with inverse document frequency and length normalization.

C. Similarity Coefficients

The associative coefficients determine the similarity in vector space model and these associative coefficients are dependent on the inner product of the document vector and query vector, and the similarity is indicated by the word overlap. The inner product is usually normalized. The most popular similarity measure is the cosine coefficient, which measures the angle between the document vector and the query vector.

VI. RESEARCH ELABORATION

- This paper contains of four page-count-based similarity scores and automatically extracted lexico-syntactic patterns from text snippets.
- Most web search engines provide Page counts and text snippets which are the main source of information.

Few problems that may occur with Page counts are:

- Page count perusal overlook the position of a word in a page
- Two words appear in a page, they might not be related with each other
- Polysemous word (a word with multiple senses). For example:
  - *apple* as a fruit
  - *apple* as a computer

Lexico-syntactic patterns

The various semantic relations *also known as,*

- *is a,*
- *part of,*
- *is an example of*

Page-count-based Similarity Scores (co-occurrence measures)

$$\text{WebJaccard}(P,Q) = \begin{cases} 0 & \text{if } H(P\cap Q) \le c \\ \frac{H(P\cap Q)}{H(P)+H(Q)-H(P\cap Q)} & \text{otherwise.} \end{cases} \quad (1)$$

$$\text{WebOverlap}(P,Q) = \begin{cases} 0 & \text{if } H(P\cap Q) \le c \\ \frac{H(P\cap Q)}{\min(H(P),H(Q))} & \text{otherwise.} \end{cases}$$

$$\text{WebDice}(P,Q) = \begin{cases} 0 & \text{if } H(P\cap Q) \le c \\ \frac{2H(P\cap Q)}{H(P)+H(Q)} & \text{otherwise.} \end{cases} \quad (3)$$

$$\text{WebPMI}(P,Q) = \begin{cases} 0 & \text{if } H(P\cap Q) \le c \\ \log_2\left(\frac{\frac{H(P\cap Q)}{N}}{\frac{H(P)}{N}\frac{H(Q)}{N}}\right) & \text{otherwise.} \end{cases} \quad (4)$$

VII. ALGORITHM FOR EXTRACTING PATTERNS

Given a set *S* of synonymous

Algorithm 3.1: EXTRACTPATTERNS(S)

comment: Given a set S of word-pairs, extract patterns.

for each word-pair $(A,B) \in S$
  do $D \leftarrow$ GetSnippets("A B")
$N \leftarrow null$
for each snippet $d \in D$
  do $N \leftarrow N +$ GetNgrams(d,A,B)
$Pats \leftarrow$ CountFreq(N)
return $(Pats)$

Figure.4: Pattern extracts from text snippets

  - n-grams : n=2,3,4, and 5

- A set *S* of closely associated word-pairs
  - 5000 word pairs of closely associated nouns from Word Net
  - 4,562,471 unique patterns
  - 80% occur less than 10 times
- A set of non-associated word-pairs
  - 5000 word pairs of non- associated nouns from Word Net
  - nouns from WordNet

Table 1: Contingency table

	v	other than v	All
Freq. in snippets for synonymous word pairs	p_v	$P-p_v$	P
Freq. in snippets for non-synonymous word pairs	n_v	$N-n_v$	N

$$\chi^2 = \frac{(P+N)(p_v(N-n_v)-n_v(P-p_v))^2}{PN(p_v+n_v)(P+N-p_v-n_v)}. \quad (5)$$

Integrating Patterns and Page Counts

Algorithm 3.2: GETFEATUREVECTOR(A, B)

comment: Given a word-pair A, B get its feature vector F.

$D \leftarrow$ GetSnippets("A B")
$N \leftarrow null$
for each snippet$d \in D$
  do $N \leftarrow N +$ GetNgrams(d, A, B)
$SelPats \leftarrow$ SelectPatterns$(N, GoodPats)$
$PF \leftarrow$ Normalize$(SelPats)$
$F \leftarrow [PF, WebJaccard, WebOverlap, WebDice, WebPMI]$
return (F)

Figure. 5: Integrating patterns and page counts

VIII. EXPERIMENTAL RESULTS

- WebOverlap (rank=18,weight=2.45)
- Web-Jaccard (rar weight=0.618)
- WebPMI (rank=138,weight=0.0001)

Table 2: Features with the highest SVM linear kernel weights

feature	χ^2	SVM weight
WebDice	N/A	8.19
X/Y	33459	7.53
X, Y :	4089	6.00
X or Y	3574	5.83
X Y for	1089	4.49
X , the Y	1784	2.99
with X (Y	1819	2.85
X=Y	2215	2.74
X and Y are	1343	2.67
X of Y	2472	2.56

Table3: Semantic Similarity of Human Ratings and Baselines on Miller-Charles' Dataset

Word Pair	Miller-Charles'	Web Jaccard	Web Dice	Web Overlap	Web PMI	Sahami [36]	CODC [6]	Proposed SemSim
cord-smile	0.13	0.102	0.108	0.036	0.207	0.090	0	0
rooster-voyage	0.08	0.011	0.012	0.021	0.228	0.197	0	0.017
noon-string	0.08	0.126	0.133	0.060	0.101	0.082	0	0.018
glass-magician	0.11	0.117	0.124	0.408	0.598	0.143	0	0.180
monk-slave	0.55	0.181	0.191	0.067	0.610	0.095	0	0.375
coast-forest	0.42	0.862	0.870	0.310	0.417	0.248	0	0.405
monk-oracle	1.1	0.016	0.017	0.023	0	0.045	0	0.328
lad-wizard	0.42	0.072	0.077	0.070	0.426	0.149	0	0.220
forest-graveyard	0.84	0.068	0.072	0.246	0.494	0	0	0.547
food-rooster	0.89	0.012	0.013	0.425	0.207	0.075	0	0.060
coast-hill	0.87	0.963	0.965	0.279	0.350	0.293	0	0.874
car-journey	1.16	0.444	0.460	0.378	0.204	0.189	0.290	0.286
crane-implement	1.68	0.071	0.076	0.119	0.193	0.152	0	0.133
brother-lad	1.66	0.189	0.199	0.369	0.644	0.236	0.379	0.344
bird-crane	2.97	0.235	0.247	0.226	0.515	0.223	0	0.879
bird-cock	3.05	0.153	0.162	0.162	0.428	0.058	0.502	0.593
food-fruit	3.08	0.753	0.765	1	0.448	0.181	0.338	0.998
brother-monk	2.82	0.261	0.274	0.340	0.622	0.267	0.547	0.377
asylum-madhouse	3.61	0.024	0.025	0.102	0.813	0.212	0	0.773
furnace-stove	3.11	0.401	0.417	0.118	1	0.310	0.928	0.889
magician-wizard	3.5	0.295	0.309	0.383	0.863	0.233	0.671	1
journey-voyage	3.84	0.415	0.431	0.182	0.467	0.524	0.417	0.996
coast-shore	3.7	0.786	0.796	0.521	0.561	0.381	0.518	0.945
implement-tool	2.95	1	1	0.517	0.296	0.419	0.419	0.684
boy-lad	3.76	0.186	0.196	0.601	0.631	0.471	0	0.974
automobile-car	3.92	0.654	0.668	0.834	0.427	1	0.686	0.980
midday-noon	3.42	0.106	0.112	0.135	0.586	0.289	0.856	0.819
gem-jewel	3.84	0.295	0.309	0.094	0.687	0.211	1	0.686
Correlation	1	0.259	0.267	0.382	0.548	0.579	0.693	0.834

IX. CONCLUSIONS

In this work, the authors discussed the problem of semantic similarity measure for words based on both page counts and text snippets which are extracted from a web. Four different word co-occurrence measures were computed using page counts.

Lexical pattern extraction algorithm is proposed that can extract various semantic relations that exist between two words. Moreover, a sequential pattern clustering algorithm is also proposed in order to identify different lexical patterns that describe the same logical or meaningful relation. To define similarity features for a word pair, both page counts-based co-occurrence measures and lexical pattern clusters were used. A two-class Support Vector Machine is used for the features that extracted for synonymous and non-synonymous word pairs that are selected from WordNet synsets. Experimental results on three referenced or benchmark data sets shows that the proposed method outperforms various baselines as well as previously proposed web-based semantic similarity measures, achieving a high correlation with human ratings. Moreover, the proposed method improved the F-score in a community mining.

REFERENCES

[1] A. Kilgarriff, "Googleology Is Bad Science," Computational Linguistics, vol. 33, pp. 147-151, 2007.

[2] M. Sahami and T. Heilman, "A Web-Based Kernel Function for Measuring the Similarity of Short Text Snippets," Proc. 15th Int'l World Wide Web Conf., 2006.

[3] D. Bollegala, Y. Matsuo, and M. Ishizuka, "Disambiguating Personal Names on the Web Using Automatically Extracted Key Phrases," Proc. 17th European Conf. Artificial Intelligence, pp. 553 557, 2006.

[4] H. Chen, M. Lin, and Y. Wei, "Novel Association Measures Using Web Search with Double Checking," Proc. 21st Int'l Conf. Computational Linguistics and 44th Ann. Meeting of the Assoc. for Computational Linguistics (COLING/ACL '06), pp. 1009-1016, 2006.

[5] M. Hearst, "Automatic Acquisition of Hyponyms from Large Text Corpora," Proc. 14th Conf. Computational Linguistics (COLING), pp. 539-545, 1992.

[6] M. Pasca, D. Lin, J. Bigham, A. Lifchits, and A. Jain, "Organizing and Searching the World Wide Web of Facts - Step One: The One Million Fact Extraction Challenge," Proc. Nat'l Conf. Artificial Intelligence (AAAI '06), 2006.

[7] R. Rada, H. Mili, E. Bichnell, and M. Blettner, "Development and Application of a Metric on Semantic Nets," IEEE Trans. Systems, Man and Cybernetics, vol. 19, no. 1, pp. 17-30, Jan./Feb. 1989.

[8] P. Resnik, "Using Information Content to Evaluate Semantic Similarity in a Taxonomy," Proc. 14th Int'l Joint Conf. Aritificial Intelligence, 1995.

[9] D. Mclean, Y. Li, and Z.A. Bandar, "An Approach for Measuring Semantic Similarity between Words Using Multiple Information Sources," IEEE Trans. Knowledge and Data Eng., vol. 15, no. 4, pp. 871-882, July/Aug. 2003.

[10] G. Miller and W. Charles, "Contextual Correlates of Semantic Similarity," Language and Cognitive Processes, vol. 6, no. 1, pp. 1-28, 1998.

[11] D. Lin, "An Information-Theoretic Definition of Similarity," Proc. 15th Int'l Conf. Machine Learning (ICML), pp. 296-304, 1998.

[12] R. Cilibrasi and P. Vitanyi, "The Google Similarity Distance," IEEE Trans. Knowledge and Data Eng., vol. 19, no. 3, pp. 370-383, Mar. 2007.

[13] M. Li, X. Chen, X. Li, B. Ma, and P. Vitanyi, "The Similarity Metric," IEEE Trans. Information Theory, vol. 50, no. 12, pp. 3250-3264, Dec. 2004.

[14] P. Resnik, "Semantic Similarity in a Taxonomy: An Information Based Measure and Its Application to Problems of Ambiguity in Natural Language," J. Artificial Intelligence Research, vol. 11, pp. 95- 130, 1999.

[15] R. Rosenfield, "A Maximum Entropy Approach to Adaptive Statistical Modelling," Computer Speech and Language, vol. 10, pp. 187-228, 1996.

[16] D. Lin, "Automatic Retrieval and Clustering of Similar Words," Proc. 17th Int'l Conf. Computational Linguistics (COLING), pp. 768 774, 1998.

[17] J. Curran, "Ensemble Methods for Automatic Thesaurus Extraction," Proc. ACL-02 Conf. Empirical Methods in Natural Language Processing (EMNLP), 2002.

[18] C. Buckley, G. Salton, J. Allan, and A. Singhal, "Automatic Query Expansion Using Smart: Trec 3," Proc. Third Text REtreival Conf., pp. 69-80, 1994.

[19] V. Vapnik, Statistical Learning Theory. Wiley, 1998.

[20] K. Church and P. Hanks, "Word Association Norms, Mutual Information and Lexicography," Computational Linguistics, vol. 16, pp. 22-29, 1991.

Measurement of Ultrasonic Velocity and Attenuation Coefficient on Cell Density and Disruption of Aspergillus Niger

Srinivasa Reddy M[1], Linga Reddy D[2]
[1]Department of Physics, Anurag Group of Institutions, Venkatapur, Hyderabad, India
[2]Deptartment of Physics, UCS, Osmania University, Hyderabad, India.
E mail : srmainampati@gmail.com, dlreddy_phy@yahoo.com

Abstract—**The present study envisaged the measurement of ultrasonic velocity and attenuation coefficient on cell density of A.niger by using novel Pulse Echo Overlap (PEO) technique. The author reported parameters like acoustic impedance, adiabatic compressibility and free length were computed by applying Jacobson's Free Length Theory (FLT). The cell density was carried out by using hemocytometer. The ultrasound transducer of 10 mm gap between transmitting and reflecting surfaces was immersed into the sonication vessels which were filled with varied cell density of 10^1 to 10^8 per ml to measure ultrasonic parameters for 2 and 10 MHz ultrasound. The sonication vessel kept inside the water bath to maintain constant environmental temperature of 303K by using automatic temperature controlled bath. In addition, the cell disruptions were carried out at 2 and 10 MHz high power ultrasound of output power of 600 watts by using ultrasonic processor. The ultrasonic velocity and attenuation coefficient were measured simultaneously with disruption of cell density as a function of sonication time. The uncertainty in measuring of ultrasonic velocity (v) is ± 0.02% and the attenuation coefficient (α) is ± 0.0015%. It was observed that both ultrasonic velocity and attenuation coefficient linearly increased with increasing concentration of cells. It was reported that the increased ultrasonic velocity was due to decreased adiabatic compressibility and free length with increased cell density. It was investigated that the attenuation coefficient was high at 10 MHz than 2 MHz high absorption losses occurred at 10 MHz. It was concluded that ultrasound waves propagated through culture of A.niger disturb the equilibrium between cells presented in the suspension and this retarded equilibrium shifts between the number of cells and their sizes and shapes, in turn, the ultrasound velocity and absorption.**

Keywords: **Pulse Echo Overlap (PEO) technique, Ultrasound, A.niger, ultrasonic velocity and attenuation coefficient**

I. INTRODUCTION

In the recent years, the measurement of ultrasonic velocity and attenuation coefficient has been adequately employed in understanding biological effects induced by ultrasound in macromolecules, microorganisms, tissues, organs and bones etc. Beside widespread well established studies of ultrasonic velocity and attenuation coefficient of biological materials in basic research, more recently the adiabatic compressibility and intermolecular free length has also been considered as an indicator with high potential in the advanced medical diagnoses and as a process control parameter in the food industry [1-3].

The lot of research has been carried out by researchers dealt with disruption and inactivation of microbes depending upon the power, pressure and frequency applied through a number of physical, mechanical and chemical effects arising from acoustic cavitations [4-6].

Resistance of different species to ultrasound differed widely which was because specific effect of ultrasound on the cell wall and differences in the cell wall structures among species [7]. Bacterial spores were much more resistant than vegetative ones and fungi were more resistant in general than vegetative bacteria [8]. Since molds and yeasts were in general more resistant to high intensity ultrasound [9], and not enough information about mold spores was available.

It has been reported that the velocity of an ultrasound wave through a medium varies with the physical properties of the medium. Liquids exhibit ultrasound velocities intermediate between those in gases and solids. With the notable exceptions of lung and bone, biologic tissues yield velocities roughly similar to the velocity of ultrasound in liquids. In different media, changes in velocity were reflected in changes in wavelength of the ultrasound waves, with the frequency remaining relatively constant [10].

The velocity of sound was determined by the density (ρ) and compressibility (K) of the medium. In a given volume the amount of material was equals to density and compressibility was a measure of how much a substance could be compacted for a given pressure. The velocity of sound in a medium can be determined by the equation

$$v = (K\rho)^{-½}$$

Where v is the speed of sound, K is the compressibility, and ρ (rho) is the density.

The various forms of interaction between ultrasonic waves and particles were necessary to understand how ultrasound can be used to characterize microbes [11-12]. The four most important types of interaction between an ultrasonic waves and a colloidal dispersion of microbes were intrinsic absorption, visco-inertial dissipation losses, thermal dissipation losses and scattering losses.

The colloidalsuspension consists of large number of individual components. The phase of each component absorbs ultrasound as a result of classical(viscous and thermal) and relaxation loss mechanisms. These losses were determined by the composition ofthe suspension, rather than by its microstructure.

Especially, there were two major classes of attenuation mechanisms were observed in ultrasonic materials characterization. The primary, absorption converts acoustic energy into heat via viscosity, relaxation, heat conduction, elastic hysteresis, etc. The absorbed energy was irreversibly lost from the acoustic field since it was dissipated in the medium. The secondary, scattering converts the energy of the coherent, collimated beam into incoherent, divergent waves as a result of wave interaction with in-homogeneities in the material.

To a first approximation, the overall attenuation coefficient of a colloidal suspension of microbes can be considered to be the sum of these various contributions (although in reality some of these mechanisms were coupled to each another). In most suspensions, one or two of the above mechanisms usually dominate the overall attenuation in a particular frequency range.

Aspergillus niger was chosenas representative species. In terms of ultrasonic measurements more research is necessary to understand their use to estimate cell density and disruptive applications in food industry,especially for fungi. The purpose of this research was aimed to measure density, ultrasonic velocity, attenuation coefficient, and evaluatescompressibility, free length and acoustical impedance as a function of cell density and sonication time.

II. MATERIALS AND METHODS

A. Sample Collection

The samples were isolated from the effected rice seed; rice bran, wheat bran and cheese were collected in sterilized screw capped bottles and transported to the Advanced Microbiology Laboratory, University College of Technology. The samples were stored under refrigeration conditions for conducting different experiments.

B. Sample Preparation

The water suspensions used for isolation were prepared by dipping the samples in pre-sterilized distilled water.

C. Media Preparation

The selected media such as potato dextrose agar (PDA) and potato dextrose broth (PDB), used for fungal development were prepared according to the methods suggested by Harrigan (1998). The pH of media was adjusted by using 0.1N NaOH and 0.1N HCl.

i. Potato Dextrose Broth

The 1 L of distilled water was used to boil diced potatoes for 1 hour and then filtered through muslin cloth. The volume of filtrate was made upto 1000 mL and then glucose was added. The sterilization of medium was done by autoclaving.

Composition of potato dextrose broth	
Ingredients	Quantity
Potatoes peeled and diced into small pieces	200g
Glucose	20g
Distilled water	1000 mL

ii. Potato Dextrose Agar

The potato dextrose broth (PDB) was converted into potato dextrose agar (PDA) by adding 1.5% agar-agar and then sterilized by autoclaving.

D. Inoculation and Incubation

The samples (water suspensions) were first inoculated on to the PDA and incubated at 30°C for 72-96 hours and growth pattern was studied according to the suggestions and methods of Harrigan (1998). The selected colonies from PDA were further transferred to PDB for growth of cells.

E. Identification of Aspergillus Niger

Aspergillus niger was identified on the basis of morphology and growth pattern according to the methods recommend by Harrigan (1998). It was based on generalexamination of growth pattern of mycelia and spores under microscope after staining.

F. Ultrasonic Measurement Studies

Firstly, the 10 mL potato dextrose broth (without inoculum) was poured into the sonication vessel (control)(15 ml glass bottles, internal diameter 21mm, flat base, 2.5 mm wall thickness). Next, the low power (2 mW) ultrasonic transducer of 10mm gap between transmitting and reflecting surfaces was immersed into the vessel. The other end of the transducer is connected to the ULTRASONIX 4400M (fabricated and supplied by Roop Telsonic Ultrasonix Limited, Mumbai.) by BNC cable and Pulse Echo pattern was observed on the CRO screen which was used to measure delay time and attenuation. Further the selected cells of A.niger from PDA inoculated into the 10 mL PDB which were poured into the sonication vessels with cell density ranging from 10^1 to 10^8 cells/ml insteps of 10^1 increments. For each cell density ultrasonic velocity and attenuation coefficient measurements were repeated for six times and average values were reported in tables. The sonication vessel kept inside the water bath upto its neck and the temperature of the water bath was maintained at constant temperature of 303K with aid of automatic temperature control unit.

G. Ultrasonic Disruptive Studies

For the ultrasound irradiation purpose the control samples (without sonication) of A.niger cultures were selected with 10^7 cell/ml.PDB inoculated 10^7 cell/ml took into the different number of sonication vessels (15 ml glass bottles, internal diameter 21mm, flat base, 2.5 mm wall thickness to a depth of 25 mm) of 10 ml each exposed to ultrasound for a period of 1 min to maximum of 6 min at 2 MHz and 1 min to maximum of 8 min at 10 MHz by using ultrasonic-processor (model USG – 600) series and an output power of 600 W i.e. 60 W/ml. Throughout the study a sonotrode of 12 mm diameter of titanium probe set at 5 mm below the surface of the culture was used. After sonication, the densities of viable cells were counted by using hemocytometer. The measurements were repeated at each sonication minute for six times and average values were reported in the tables.

H. Ultrasonic Velocity Measurements

The novel Pulse Echo Overlap (PEO) technique was introduced to measure ultrasonic parameters in PDB of A.niger. The ultrasonic transducer was immersed in PDB. It generates a pulse of ultrasound which travelled across the sample, was reflected from the bottom wall of the measurement cell, travels back through the sample, and was then detected by the same transducer.The single transducer was used to both transmit and receive ultrasonic pulses. The ultrasound velocity in the medium was found from the measured delay time difference (Δt) and earlier found length of the measurement chamber (d).

$$v = \frac{2d}{\Delta t}\ cm/sec \qquad \text{------- (1)}$$

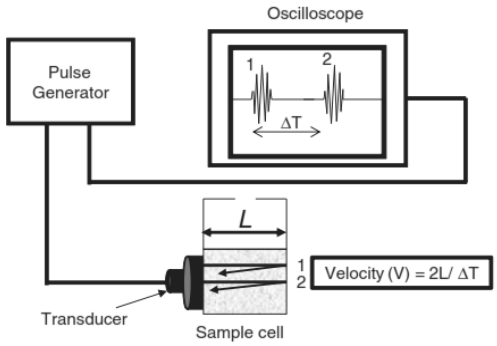


Figure 1. Schematic representation of Ultrasonic velocity measurements

The uncertainty in velocity measurements by this technique were ± 0.02%.

I. Attenuation Coefficient Measurements

The multiple reflections of ultrasound pulses in the sample produced consecutive echoes which were observed on CRO screen. The attenuation coefficients of the fungal cultures were carried out by measuring the amplitudes of transmitted pulses of selected two successive echoes on CRO screen. The ultrasonic attenuation coefficient was calculated by using the following formula.

$$\alpha = \frac{1}{2l}\ln\left(\frac{A_n}{A_{n+1}}\right) \qquad \text{nep/cm} \quad \text{----- (2)}$$

Where 2*l*was distance travelled and A_n / $A_{n+1}$was the ratio between two successive echoes of A_n and A_{n+1}.The uncertainty in attenuation coefficient **(α)** measurements were ±0.0015%.

J. Measurement of Cell Density

The fungal cells were counted with the help of Hemocytometer. The microscope at 100x magnification was used to view the spores. Under the microscope, grids of 9 squares were identified. After that, focus the microscope on one of the 4 outer squares in the grid. The each square should contain 16 smaller squares. Then, Count all the spores in the four 1 mm corner squares. If there are too many or few spores to count, repeat the procedure either concentrating or diluting the original suspension as appropriate.

Viable cells/ml = Average viable cells count per square x Dilution Factor.

K. Density Measurements

The bicapillary Pyknometer of 10 ml volume was used to find density of bacteria cultures. The micro pipette was adapted to transfer A.niger cultures to PDB in the laminar flow chamber to avoid contamination with air and body.The density of bacteria cultures was measured by using the following procedure

Mass of the empty bicapillary pyknometer = $w_1$gm
Mass of the bacteria culture + pyknometer= w_2 gm
Mass of the culture (m) = $w_2 - w_1$ gm
Volume of the culture = V cm^3

Density of the bacteria culture (ρ)

$$= \frac{Mass\ \ of\ \ the\ \ culture\ (m)}{Volume\ \ of\ the\ \ culture\ (V)}\ \ gm/cm^3 \qquad \text{------ (4)}$$

The accuracy in measuring density of bacteria cultures was 2 parts in 10^5.The single pan electrical balance was used to find the masses of the samples. The accuracy of measuring the masses was±0.01mg**.**

L. Computed Parameters

i. Compressibility (βs)

The compressibility can be calculated as

$\beta_s = 1/\rho v^2$ cm^2/dynes -------- (5)

Hereρwas the density and was the ultrasonic velocity

ii. Free Length (Lf)

The free length wascalculated by substituting compressibility in below equation

$L_f = K \frac{1}{\beta_s}$ A.U. ------- (6)

Where K was Jacobson's temperature constant was = 631 x 10^{-6} at 303 K

iii. Acoustical Impedance (z)

The equation to calculate acoustical impedance is

$Z = \rho v$ ---------- (7)

Where ρwas the density and νwas the ultrasonic velocity

III. RESULT AND DISCUSSIONS

The variation of density and ultrasonic parameters with cell density were reported in the table 1. Figure 2 showed that density exponentially increased and ultrasonic velocity linearly increased with increasing cell density. The adiabatic compressibility and free length decreased exponentially with increasing cell density as shown in figure 3. It was observed that as the number of cells increasing in culture might be the reason to increase density. As we know that free length between the cells might be decreased with increased density of culture and in turn caused to decrease compressibility. It was investigated that the decreased compressibility of culture with increased

cell density could be the primary reason for the increased ultrasonic velocity.

Figure 4 showed that the variation of attenuation coefficient at 2 and 10 MHz with increasing cell density was followed the linear path, because at low cell concentration the attenuation coefficient was only due to absorption losses, as the concentration of cells increased in the suspension, the scattering losses were initiated due to availability of more and more scatters. It was reported that the visco-internal and thermal loss mechanisms usually dominateat nearly low frequencies, but the intrinsic absorption and scattering losses usually dominateat relatively higher frequencies. Usually absorption losses were directly proportional to square of frequency in liquids due to this attenuation was reported more at 10 MHz than 2 MHz for A.niger. Therefore, in general attenuation coefficient at 10 MHz was reported greater than 2 MHz. It was observed that the variation of attenuation coefficient was more happened at 2 MHz than 10 MHz because scattering losses were inversely proportional to cell size of A.niger.

The density and ultrasonic parameters variations as a function of sonication time were reported in table 2. It was observed that density suddenly fell after few minutes of sonication time because cell density in PDB rapidly decreased with sonication time (Figure 5). Figure 6 showed that the ultrasonic velocity linearly decreased with increased sonication time. It was found that the sonication time required to bring the ultrasonic velocity to minimum value was less at 2 MHz than 10 MHz because the cavitation effect was more at 2 MHz than 10 MHz. It was investigated that the formation of transient cavitation occurred early at 2 MHz than 10 MHz which was held responsible for cell disruption. The variation of ultrasonic velocity with sonication time was also confirmed linearly increased compressibility and free length of suspension with sonication time.

It was observed that attenuation coefficients were almost linearly decreased with sonication time (Figure 9). It was found that with increased sonication time the numbers of cells present in PDB were decreased linearly (Figure 10). It was reported that the scattering losses were decreased with reduced numbers of cells present in PDB and hence attenuation coefficient.

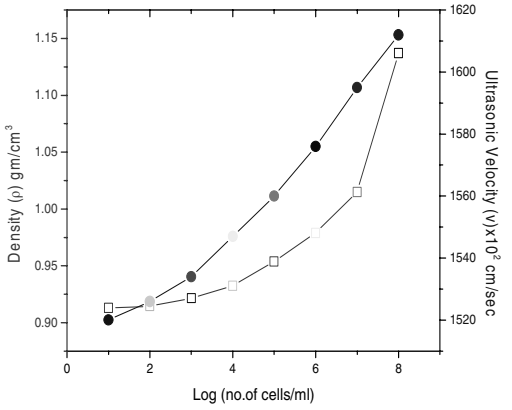


Figure 2. Graphic representation of density and ultrasonic velocity with cell density

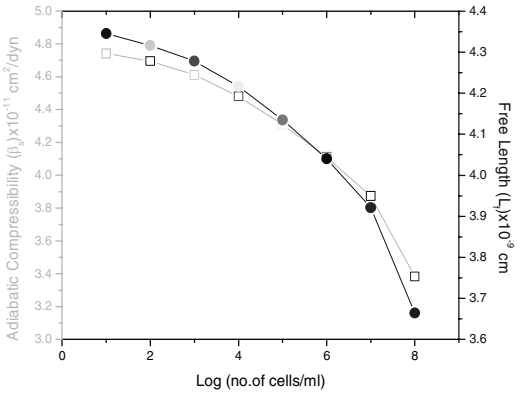


Figure 3. Graphic presentation of adiabatic compressibility and free length as a function of cell density

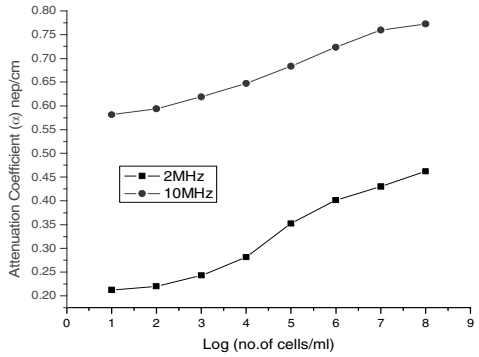


Figure 4. Comparison of attenuation coefficient as function of cell density

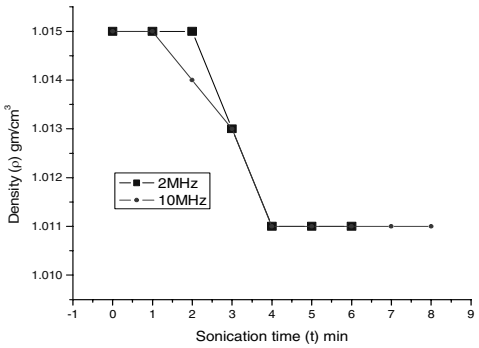


Figure 5. Comparison of density as a function of sonication time

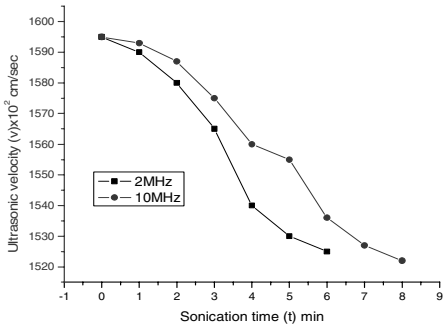


Figure 6. Comparison of ultrasonic velocity as a function of sonication time

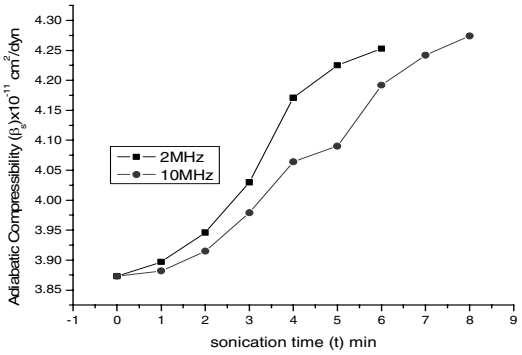


Figure 7. Comparison of adiabatic compressibility as a function of sonication time

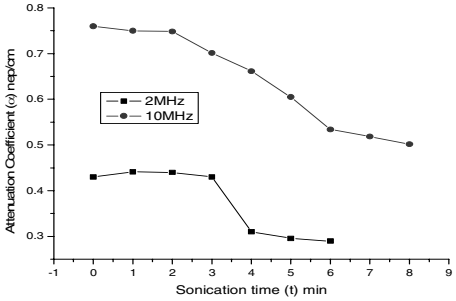


Figure 9. Comparison of attenuation coefficient as a function of sonication time

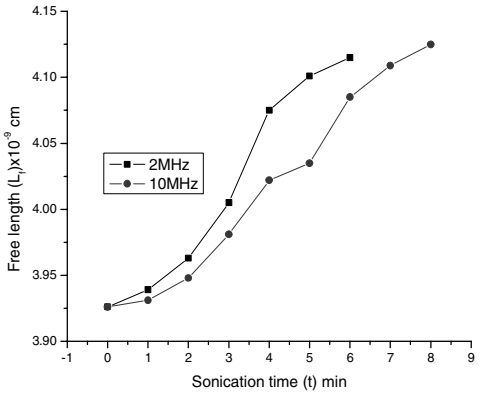


Figure 8. Comparison of free length as function of sonication time

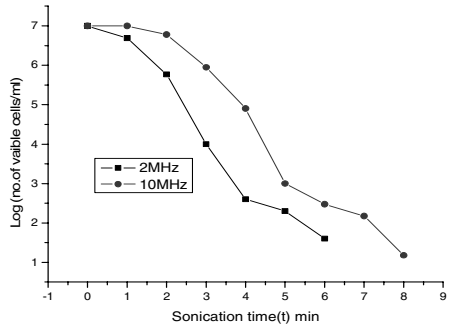


Figure 10. Comparison of no. of viable cells as a function of sonication time

Table1: The Number of spores per ml, Density, Ultrasonic velocity, Adiabatic compressibility, free length, Attenuation coefficient and Acoustical Impedance of Aspergillus Niger at 303K.

No.of cells/ml	Density (ρ) (gm/cm^3)	Ultrasonic Velocity (v) × 10^{2}cm/sec	Adia.Comp (β_s) X 10^{-11} cm^2/dyn	Free length (L_f) X 10^{-9} cm	Attenuation Coefficient (α) nep/cm		A. I. (z)×10^5 gm/cm^2-sec
					2 MHz	10MHz	
10^1	0.9129	1520	4.741	4.345	0.2125	0.5815	1.387
10^2	0.9145	1526	4.695	4.316	0.2200	0.5941	1.395
10^3	0.9216	1534	4.611	4.278	0.2431	0.6189	1.413
10^4	0.9324	1547	4.481	4.217	0.2815	0.6472	1.442
10^5	0.9538	1560	4.308	4.135	0.3524	0.6834	1.487
10^6	0.9789	1576	4.113	4.040	0.4013	0.7236	1.542
10^7	1.015	1595	3.873	3.921	0.4301	0.7597	1.618
10^8	1.137	1612	3.384	3.664	0.4621	0.7723	1.832

Table 2: Sonication time, Density, Ultrasonic velocity, Adiabatic compressibility, free length, Attenuation coefficient, Acoustical Impedance and Number of viable spores per ml of A.niger

Sonication time (t) min.	Density (gm/cm3)	Ultrasonic Velocity (v) × 102cm/sec	Adia.Comp (βs) X 10-11cm2/dyn	Free length (Lf)X 10-9 cm	Attenuation Coefficientnep/cm	A.I.(z)× 105 gm/cm2-sec	Number of viable cells/ml
0	1.015	1595	3.873	3.926	0.4301/0.7597	1.618	107
2 MHz							
1	1.015	1590	3.897	3.939	0.4412	1.613	5x106
2	1.015	1580	3.946	3.963	0.4400	1.603	6x105
3	1.013	1565	4.03	4.005	0.4302	1.585	104
4	1.011	1540	4.171	4.075	0.3100	1.556	400
5	1.011	1530	4.225	4.101	0.2956	1.546	200
6	1.011	1525	4.253	4.115	0.2898	1.541	40
10 MHz							
1	1.015	1593	3.882	3.931	0.75	1.616	107
2	1.014	1587	3.915	3.948	0.7485	1.609	6x106
3	1.013	1575	3.979	3.981	0.7015	1.595	9x105
4	1.011	1560	4.064	4.022	0.6615	1.577	8x104
5	1.011	1555	4.09	4.035	0.605	1.572	1000
6	1.011	1536	4.192	4.085	0.5341	1.552	300
7	1.011	1527	4.242	4.109	0.5187	1.543	150
8	1.011	1522	4.274	4.125	0.5019	1.537	15

IV. CONCLUSIONS

It was concluded that the ultrasonic velocity and attenuation coefficient strongly depends on cell size and number of cell present in PDB. It was hypothesized that scattering losses were occurred more at 2 MHz than 10 MHz for A.niger. It was learnt that the data of ultrasonic parameters could be used to estimate the fungal cell density. It was also learnt that the variation of ultrasonic parameters with sonication time could be used for image production.

REFERENCES

[1] Srinivasa Reddy. M and Linga Reddy. D, "Ultrasonic dosimetry studies in microbial growth periods – synergies of 2 MHz and 10 MHz ultrasonic frequencies," *IntInt. Journal of Research in Biochemistry and Biophysics,Vol.**3**(3) , p 21-28*, 2013.

[2] Koward.W.Nowak et al., "A comparison of methods for the determination of sound velocity in biological materials: A case study", *Ultrasonics*, Vol. **53**, pp. 923-927, 2013.

[3] T.S. Awad and H.A. Moharram , et.al. , "Applications of ultrasound in analysis, processingand quality control of food:A review," *Food Research International,*vol. 48, pp. 410-427, 2012.

[4] William. G and Aaron Ross. S, " Ultrasound increases the rate of Bacterial cell growth," *BiotechnolProg*, vol. 19(3), pp. 1038-1044, 2003.

[5] K. Sathi Reddy, Y.HariBabu, M. K. Tandle and D.Linga Reddy, "Ultrasonic Studies on the different veterinary Samples," *J. Bioscience, Biotechnology Research Asia*, vol. 4 (2), pp. 777-780, 2007.

[6] Srinivasa Reddy. M and Linga Reddy. D, "Investigation of Ultrasonic Parameters on Disruption of *S.epidermidis* at Ultrasound exposure of 20 kHz and 36 kHz frequencies," *Proceedings of 20thNational Symposium on Ultrasonics (NSU – XX),* ISBN 978-81-926579-1-2, pp 88-93, 2014.

[7] Earnshaw, R. G. , " Ultrasound a new opportunity for foodpreservation", *Ultrasoundin food processing,* pp. 183–192, 1998.

[8] Alliger, H., " Ultrasonic disruption", *American Laboratory*, vol.10, pp.75–85, 1975.

[9] Guerrero, S., Lopez-Malo, A., &Alzamora, S. M., " Effect of ultrasound on the survival of Saccharomyces cerevisiae: influence oftemperature, pH and amplitude" *Innovative Food Science andEmerging Technologies*, vol. 2, pp.31–39, 2001.

[10] Voleišiene, A. Voleišis, "Ultrasound velocity measurements in liquid media" *ULTRAGARSAS*

.

(ULTRASOUND), Vol. 63 (4), pp 7-19, 2008.

[11] U. Kaatze, B.K. Lautscham, `Acoustic Absorption Spectroscopy of Liquids between 0.15 and 3000 MHz: II. Hypersonic ComparatorTechnique', *J. Phys. E.*, **21**, 402±406, 1988.

[12] Kertzfield, K., and Litovitz, T. *Absorption and Dispersion of Ultrasonic Waves*. New York, Academic Press, 1959.

[13] N. Bilaniuk and G. S. K. Wong, *J. Acoust. Soc. Am.*, vol. 93, pp.1609–12, 1993.

A Bird's Eye View of Census 2011

P Sai Shankar
Department of H&S, C.V.R College of Engineering, Ibrahimpatan, Hyderabad, India
E mail Id: saishankar26@gmail.com

Abstract— **The first census was held in India during 1872. Since then, Census in India has been held uninterruptedly for every ten years. The 2011 census of India is the 15th Census and the 7th after independence. The Census plays a vital role in the study of - Demography, 'Literacy & Education', 'Fertility and Mortality', 'Scheduled Castes and Scheduled Tribes', Migration and Economic Activity', Language and Religion. This paper compares Census 2011 with 2001 Census in the various fields of study. According to the 2011 Census data, India has a population of 121 million. It is the second highest populous country in the world. In India the population is increasing enormously, but child sex ratio and sex ratio is at an alarming stage. Inspite of it, one encouraging factor is that Literacy rate is continuously increasing from the time of Independence.**

Index terms— **Census Data (2001 and 2011), Demography, Literacy, Economic Activity.**

I. INTRODUCTION

A population Census is the process of collecting, compiling and analyzing demographic, cultural, social, and economic data relating to all persons in the country. In our country census takes place once every ten years. For every ten years the Ministry of Home Affairs, Government of India will conduct the Census. The Indian Census has a strong history behind it. India is one of the country in the world, which has a remarkable history of conducting Census for every ten years. In the history 'Rig Veda' tells that during 800-600 BC there was some kind of Population count was maintained. In 321 BC, Kautilya's Arthasastra, focused on Census taking as a measure of State policy for purpose of taxation. During the period of Mughal king Akbar, they used to collect the data from the villages for collecting the taxes. After that a systematic and non synchronously population census were conducted between 1865 and 1872 in different parts of our country.But the first synchronously population census were conducted in 1881.

Why is Census necessary?

- *Demography (the statistical study of human population)*
- *Economic Activity*
- *Literacy & Education*
- *Urbanization*
- *Fertility & Mortality*
- *Language & Religion*
- *To make decisions regarding public welfare*
- *Monitoring ongoing scheme*
- *Plan for future*

II. HIGHLIGHTS ON 2011 CENSUS

The Slogan of the 2011 Census –*"Our Census Our Future"*

A. Administrative Units

Mr. C Chandramouli is the Registrar general and Commissioner of 2011 Census. Census 2011 has covered 640 districts, 5767 Sub-districts (Tahsil), 7742 Towns, 6.41 lakh Villages and 1.2 Billion people in India. In 2011 census forms were printed in sixteen languages. The total Cost of 2011 Census is Rs.2209 Crores. i.e., Cost per person is Rs.18. In 2001 the total Cost of Census was Rs.1403 Crores i.e., Cost per person was Rs.14. In 2011 Census the total cost was increased by 806 Crores, when compared to 2001 Census .

B. Population

According to Census 2011, the population of India is 12.10 billion consisting of 6.23 billion males and 5.865 billion female populations. Comparing to 2001, the total population is 10.287 of which 5.323 billion are males and 4.965 billion are females. Interestingly, the total population of India is more than the combined population of Pakistan, U.S.A., Japan Brazil, Bangladesh and Indonesia put together. The population of India has increased by more than 1.81 billion during the decade 2001-2011, Increase among males is 91.15 million and Increase among females is 89.95 million during the decade 2001-2011. In Census 2011, In 2011 census the highest population is recorded in Uttar Pradesh with 200 million. It was the most populous State in India. Brazil country population is less than the population of Uttar Pradesh. Maharashtra and Uttar Pradesh (312 million), is greater than the population of USA.

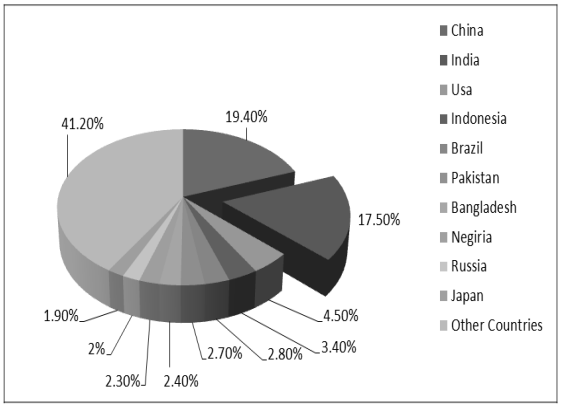


Figure 1.Population of India compare with other Countries

From the Figure 1.we can observe that highest share of population is China (19.40%) next is India (17.50%). According to the U.N.O (United Nations Organization) by the end of the year 2030 India's population will crosses the China's population.

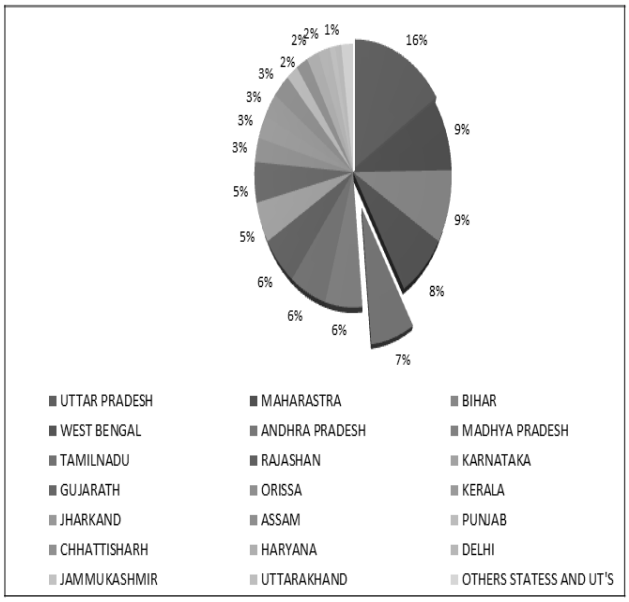


Figure 2. Population share of different states in India in 2011 census.

Figure 2. shows the highest population share is Uttar Pradesh (16%) and Andhra Pradesh is in the fifth position (7%). Here the population of Uttar Pradesh is more than the population of Brazil.

TABLE I

SMALLEST AND HIGHEST POPULATION STATES/UTS IN INDIA IN CENSUS 2011

BOTTOM 5 STATES AND UTs	
Lakshadweep(UT)	64,429
Daman & Diu(UT)	2,42,911
Dadra Haveli(UT)	3,42,853
Andaman Islands(UT)	3,79,944
Sikkim(STATE)	6,07,688

TOP 5 STATES AND UTs	
Uttar Pradesh (STATE)	199,581,477
Maharashtra(STATE)	112,372,972
Bihar(STATE)	103,804,637
West Bengal(STATE)	91,347,736
Andhra Pradesh(STATE)	84,665,533

In absolute numbers, out of the total increase of 181 million in the last decade, the contribution of rural and urban areas is equal (91.0 million each). Maharashtra has the highest share in urban population with 50.8 million (13.5% of the country's urban population) whereas Uttar Pradesh has the largest share in rural population at 155.3 million (18.6% of the country's rural population). For the first time since Independence, the increase in population is less in rural area that in urban areas.

D. Population Growth Rate

The rate of growth in the population in India during the year of 2001-11 is 17.7%. In that males growth rate is 17.19% and female growth rate is 18.12%. Hence during 2001-11 the male's growth rate is less than the female growth rate. Since Independence (1951-61), the percentage decadal growth during 2001-11 has registered the sharpest decline in the population. During 2001-11the percentage of growth rate has decreased of 3.90 percentage points from 21.54 to 17.64 percent, but among the 28 States, Bihar (25.4%) has registered highest decadal growth in population. The following bar chart represents the percentage of growth rate from 1951-2011 in India.

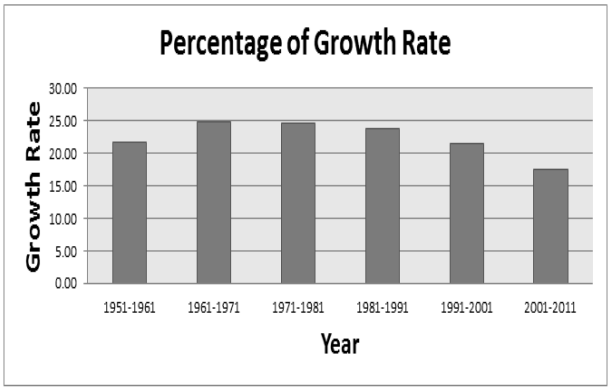


Figure 3. Growth rates from the year 1951-2011.

From the Figure 3. we can observe from 1951-81 the growth rate has been increasing. After that from 1981-2011 the growth rate is decreasing but compared to all the years, during 2001-11, the percentage of growth rate is less. In this period, 14 States/UTs have recorded population growth above 20%.

E. Population(0-6 years)

In 2011 the population (0-6 years) is 158.7 million consisting of males 82.9 million and Females 75.8 million. In 2001 Census the child population is 163.8 million (Males 85 million and Females 78.8 million). The difference in child population during the period 2001-11 is -50, 48,108 (Males -20, 56, 132 and Females -29, 91, 976). Hence we can say that child population (0-6 years) is decreased during 2001-11. Uttar Pradesh (29.7 million), Bihar (18.6 million), Maharashtra (12.8 million), Madhya Pradesh (10.5 million) and Rajasthan (10.5 million) constitute more than 50% children in the age group of 0-6 years.

F. Sex Ratio (Number of Females per 1000 Males)

The sex ratio in the country which was at 940 in 2011 census. In 2011 census Sex ratio has increased by 7 points compare to 2001. This is the highest Sex Ratio recorded since 1971 Census and a shade lower than Census 1961. In rural areas in India, the sex ratio has increased from 946 to 949. The increase in urban areas has been of 29 points from 900 to 929. In Census 2011, Kerala has recorded the highest sex ratio (1084), in Rural (1078) and the

corresponding value in Urban population (1091) respectively. The lowest sex ratio in rural areas has been recorded in Chandigarh (690) and the corresponding value in Urban areas has been recorded in Daman & Diu (551). Three major States (Jammu Kashmir, Bihar & Gujarat) have a decline in Sex Ratio in 2011 Census. In this Census the sex ratio in the Rural areas (947) has increased by one point when compare to 2001 (946), in Urban areas sex ratio (926) is rapidly increased as compared to 2001 (900). The important point we should notice is that in Urban areas has shown increase of 3.9 million and decline of 8.9 million children in Rural areas. The following line chart represents the sex ratio from 1951-2011.

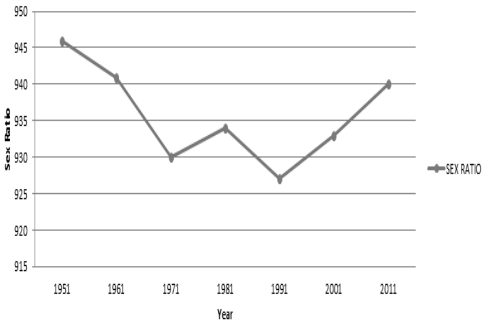


Figure 4. Sex Ratio from the years 1951-2011.

G. Child Sex Ratio (0-6 years)

In 2011 Census, child sex ratio (0-6 years) is 914. During 2001-11, child sex ratio rapidly decreased from 927 to 914 (-13 points) in the country. This is the lowest sex ratio since the year 1961. In Urban areas, the fall has been by 1 point (906 to 905) and in Rural areas, the decline has been 11 points (934 to 923) over the last decade 2001-11. In Rural areas Delhi (814) has recorded the lowest child sex ratio and Chhattisgarh (977) the highest child sex ratio during the year of 2011. In Urban areas Haryana (832) has recorded the lowest child sex ratio and Pondicherry (975) the highest child sex ratio. The following line chart represents the trends in child sex ratio from 1951-2011.

From the Figure 5.we can say that the Child Sex Ratio has been decreased from 1951-2011 Census. There will be an upward trend in the Child Sex Ratio (0-6) in Mizoram, Himachal Pradesh, Punjab, Tamil Nadu, Gujarat and Haryana,. In all remaining 27 States/UTs, the Child Sex Ratio has declining during the period 2001-11.

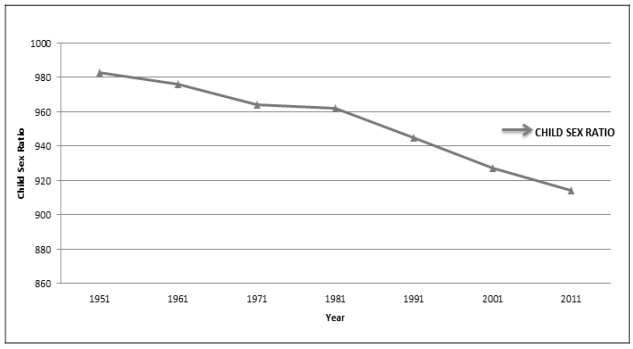


Figure 5. Growth rates from the year 1951-2011.

H. Effective Literacy Rate (Population aged 7 and above)

The Effective Literacy Rate has recorded in India in Census 2011 is 73.0%. In Urban areas literacy rate is 84.1%, and in Rural areas, literacy rate is 67.8% . There has been a positive increase of 8.2 percentage points (9.1 percentage points in rural areas and 4.2 percentage points in urban areas) during the last decade 2001 -11. In 2011 the Population (aged 7 and above) is 1051 million (Males 540.7 million and Female 510.6 million). In 2001 Population (aged 7 and above) is 864 million (Males 447.2 million and female 417.6). Figure 6 shows that the literacy rate is increasing gradually since 1951.

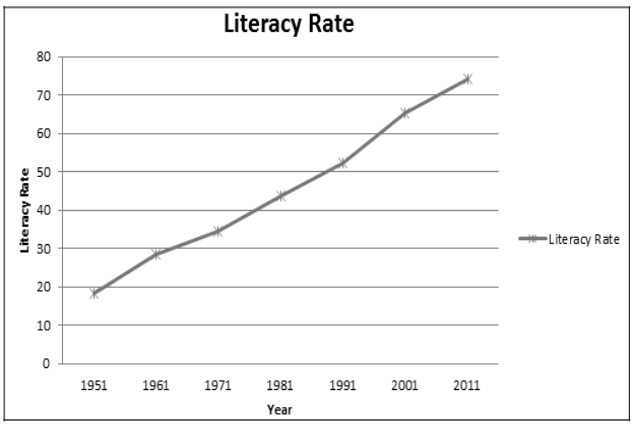


Figure 6. Litearcy rates from the year 1951-2011.

(i) Male Literacy

In 2011 Census the male Literacy rate in India is 80.9% (Rural areas 77.2% and Urban areas 88.8%). The highest male literacy rate in Rural areas has recorded in Kerala (95.4%), while the lowest is recorded from Arunachal Pradesh (67.4%). In Urban areas, the lowest male literacy rate has been recorded in Uttar Pradesh (80.4%) and the highest is in Mizoram (98.0%).

(ii) Female Literacy

In 2011 Census the female Literacy rate in India is 64.6% (Rural areas 57.9% and Urban areas 79.1%). The highest female literacy rate in Rural areas has recorded in Kerala (90.8%), while the lowest has recorded in Rajasthan (45.8%). In Urban areas, the lowest female literacy rate is returned from Jammu & Kashmir (69.0%) and the highest is in Mizoram (97.3%). A significant milestone reached in Census 2011 is that a decline of 3, 11, 96, 847 among illiterates is noted. Ten States and Union Territories(U.Ts) Lakshadweep, Puducherry , Delhi, Kerala, , Goa, Tripura, , Daman & Diu, , Chandigarh, Mizoram and Andaman & Nicobar Islands have achieved literacy rate of above 85 per cent in 2011 Census. The important factor (Literacy rate) is continuously increasing over the period from 1951 to 2011. In 2011 census has recorded the highest literacy rate (73%) from Independence.

III. CENSUS 2011 FIGURES AT GLANCE

TABLE. 2

		2001	2011			
Villages		6.38 lakh	6.41 lakh			
Towns		5161	7742			
Tehsils		5463	5767			
Districts		593	640			
States & UT		35	35			
people		1.02 Billion	1.2 Billion			
		Absolute			Percentage	
		Total	Rural	Urban	Rural	Urban
Total Population	Persons	121.0	83.3	37.7	68.8	31.1
	Males	62.3	42.7	19.5	68.5	31.3
	Females	58.6	40.5	18.1	69.1	30.8
		Absolute			Percentage	
		Total	Rural	Urban	Rural	Urban
Decadal change 2001-2011	Persons	18.1	9.04	9.08	12.18	31.8
	Males	9.1	4.6	4.52	12.12	30.06
	Females	8.9	4.4	4.57	12.25	33.73
Sex Ratio		940	947	926		
Child Population in the age group 0-6 years		Absolute			Percentage	
		Total	Rural	Urban	Rural	Urban
	Persons	15.8	11.7	4.1	14.11	10.93
	Males	8.2	6.1	2.1	14.32	11.07
	Females	7.5	5.6	1.95	13.9	10.78
Child Sex Ratio		914	919	902		
		Absolute			Percentage	
		Total	Rural	Urban	Rural	Urban
Literates	Persons	77.8	49.3	28.5	68.91	84.98
	Males	44.4	28.8	15.6	78.57	89.67
	Females	33.4	20.4	12.9	58.75	79.92

IV. CONCLUSIONS

According to the census 2011 data

- India's population is 1.21 billion. In the 2011 census the country has added 181 million new people. The good news is that 17.64%, the rate of growth during 2001-11 represents the sharpest decline since Independence.
- Encouraging news is that literacy rate in India is increasing continuously from the time of Independence. A significant milestone reached in Census 2011 is that a decline of 3,11,96,847 among illiterates is noted. And also female literacy rate is also improved over the decade. The following are schemes of Indian government for improvement of Literacy rate in India.
    1. Operation Black Board Scheme
    2. Sarva Shiksha Abhiyan
    3. Computer Literacy and study schemes in Schools.
- The biggest shock in this 2011census is the decline in the child sex ratio at 914 girls for every 1000 boys. The reasons behind these trends are:
    1. Son preference trend
    2. Sex selection practices
    3. Causes of childhood deaths
    4. Rising costs of dowry.
- Another encouraging factor is Sex Ratio. In 2011 census the sex ratio in the rural is (947) is increased one point compare to 2001 (946), but in Urban areas in 2011 sex ratio (926) is rapidly increased as compared to 2001 (900). When we compare rural and urban areas in 2011, in the urban areas sex ratio is less than rural areas. The reasons behind this trends:

    1. Medical facility (In urban areas sex selection abortions are high)
    2. Son preference and daughter discrimination
    3. Infanticide (Intentionally killing of children under age of 12 months)

REFERENCES

[1] www.mospi.com
[2] www.indiastat.com
[3] http://www.census2011.co.in/
[4] http://censusindia.gov.in/
[5] A.Malhotra and R.Pande —Son preference and Daughter neglect in India,2006
[6] Dr.Sukwanth singh - Population and Census

APPENDIX

The ratios in this paper are computed using the following formulae:

1) Population Percentage = $\frac{\text{Total population of the State}}{\text{Total Population of a Country}} \times 100$

2) Growth Rate = ((Current year population-Previous year population) ×100) ÷ (Previous year population).

3) Sex Ratio = $\frac{\text{Number of females}}{\text{Number of males}} \times 1000$

4) Child Sex Ratio in age group 0-6 year =

{((Total Number of female children in age group 0-6 years)) × 1000) ÷ (Total Number of male children in age group 0-6 years)}

5) Effective Literacy rate of population is defined as the percentage of literate persons whose age is 7 and above and who can both write and as well as read in any language has been consider as literate to the total population aged 7 and above.
Effective Literacy Rate =
((Total Number of Literate people whose age is 7 and above) × 100) ÷ (Total Population of age group 7 and above))

Innovation and its Importance for Contemporary Professional Life

(A conceptual paper with interpretation of ideas based on a close study of the existing works in the domain)

Dr. M.S.Bhat[1], Harivardhagini S[2]
[1,2]Deptartment of EIE, CVR College of Engineering, Hyderabad, India
Email: bhatms@rediffmail.com, arvind.biju@gmail.com

Abstract--**The paper provides an interpretation of work already done in the area of innovation by defining the four dimensions of innovation and their importance. The need for innovation and the kinds of innovation is discussed. The way the culture of innovations can be fostered and the importance of innovation in contemporary life is brought out clearly towards the end of the article**

Index terms-- **Dimensions of innovation, conceptual fluidity, conceptual flexibility, complexity orientation, Fostering culture of innovation, contemporary life and innovation.**

I. INTRODUCTION

Innovation is a word which strikes a pleasant and familiar chord in every one of us. For that matter, who does not like to be known as innovative person? All of us cherish this word and would willingly do our best to have this wonderful and even enviable quality of human being, becoming part of our persona. Indeed, innovation is essential for society to remain vibrant and sustain. Be it in the realm of individual life or even for a community and society, constant innovation is the life breadth. Without innovation, the society would be condemned to remain stagnant and cannot ever progress. Even when a country is not endowed with natural resources, it can quite simply, more than overcome this handicap, with collective innovative mind of its human resources and become an economic superpower, as has been amply proved by the example of Japan-- which rose like Phoenix from ashes after being completely defeated and humiliated at the end of Second World War. To an extent, so also was the case of Germany. Even when all the fortunes desert a nation leaving it in shambles, it has been proven that one can rebuild afresh everything quickly, if and only if, nation's people are resilient, resourceful and more than anything else innovative!

What is innovation? How is it different from invention? Innovation is basically a process of looking for constant improvement in any thing we do. It is a quality of mind which looks for areas where one can better others; or if one is already the best in the sphere, the desire to better oneself even further. Invention on the other hand implies—discovery-- usually of a product or process for the first time. Needless to say that Invention is just an end by itself, whereas innovation is the means of achieving this particular end. Not all innovations need to result in invention. For the outcome of inventions tend to be discrete rather than continuous to an observer making the gap between one invention and the other too far and infrequent. However, innovation is essentially a state of mind and is the basic character of human being to improve upon something existing. Innovation is more of journey than arriving at a particular destination. And, as such, innovation as a quality is more enduring and beneficial to human kind than any other entity.

II. DIMENSIONS OF INNOVATION

Innovation, thus, is a state of mind which draws heavily from the fount of creativity. Innovation is founded on creativity and creativity implies thinking our-of- box. Creativity has four clear components or ingredients according to Rastogi R.N. (1996). They are conceptual fluency, conceptual flexibility, originality and complexity orientation. Rather than viewing them as components which together constitute the 'creativity', these can be viewed more appropriately, as the basic dimensions defining creativity. Though, it may be difficult to come up with exact measures of each of these dimensions and test the relevance of these dimensions empirically using proven statistical techniques, it can be logically argued that each of the four dimensions described above are by and large, mutually independent, and are vital for creativity. And creativity encompasses these basic elements, and is significant only when each of the elements is present in adequate measures. This point would indeed be clear, when one explores, what these four terms imply, and exactly what they mean in the context of creativity.

III. CONCEPTUAL FLUENCY

Conceptual fluency refers to the level of mastery over the subject of an individual. The extent of knowledge must be to such a level, enabling the person to grasp fully all the aspects of the subject comprehensively and explain it to others neatly in clear terms in such a way that others feel fully satisfied, and for all practical purposes, consider the person as immediate and convenient source of knowledge on the subject. It is indeed quite simple to understand its importance as a dimension of creativity, as one cannot be innovative or creative without a strong streak of basic knowledge in the field where one intends to contribute. The creativity requires application of specific knowledge and the knowledge should be forthcoming unhindered and freely. For this, one needs to acquire a strong foundation of knowledge either theoritical or experiential on very firm footing. The stronger the foundation, better are the chances of creative ideas to emerge. If we examine the genesis of all earlier innovative endeavors, it would be

apparent that the new innovation is mostly a continuum of all the earlier innovation. Even where it is, sort of breakthrough, and clear break from the past, on closer examination, it would be seen that the present innovation would have made use of some or other element of the earlier work on this domain. If one cannot leverage the existing knowledge for future advancement, it is clear that very little new knowledge can be added ab initio or a fresh one without any link to the available fund of knowledge. The ability to move cognitively along one's sphere of chosen area effortlessly and freely is thus termed as conceptual fluency. Conceptual fluency is thus the basic and essential dimension of creativity. If one has difficulty in navigating or freely traversing in one's own chosen area of specialty, it is clear that very little creativity can forth come from the individual concerned. The conceptual fluency therefore, should logically enable a person to generate large number of solution for a given problem based on one's expertise in the field. In the chosen area of knowledge, the conceptual fluency would eventually enable the mind to become very fertile due to complete internalization of the concepts already imbibed, and seeds of ideas get germinated and sprouted, no sooner the problem is encountered.

IV. Conceptual Flexibility

Conceptual flexibility refers to the cognitive ability of the person to move seamlessly from one perspective to another or from one frame of reference to another. If a person has fairly good understanding and grasp of different spheres of knowledge the process would seem quite natural. The ability to cut loose from rigid boundary of one domain of knowledge and move to a different domain can be indeed a very valuable quality of a person, enabling him to synthesize something totally different and new. It is apparent that analogies abound this world as one can discern patterns in nature. And patterns do have similarities. It is always fascinating to learn that a gear-train in the domain of mechanical engineering is no different from transformers of electrical technology, a mechanical spring is analogous to inductor of electrical engineering and a flywheel is mechanical equivalent of capacitor. The analogy is not just confined to components; it extends to sub-systems and even system too. Most of the branches of social sciences extensively draw analogy from physical sciences to buttress their points. For the world is full of symmetry and pattern; and yet each one of us is singularly different to the extent of being unique! The conceptual flexibility, therefore, is a quality which enables one to liberally go for transfusion of ideas from one sphere of knowledge to another with singular purpose and yet respect the uniqueness of each sphere of knowledge. Without the development and use of analogies, and analogical tools from the 'foreign' or external disciplines, one cannot hope to develop one's own field of interest. The insight acquired through analogies is often spectacular and quite deep. The phenomenon of conceptual flexibility is akin to an individual moving freely and effortlessly across the international borders, fully respecting the sovereignty and sanctity of the international borders of the nations of the world. Conceptual flexibility is thus a quality, enabling a person to see a problem from many different angles or perspectives and then arriving an integrated picture with all its subtle nuances . In this context, it is worth recalling the saying of Oliver Williams. According to him, 'Many ideas grow better through symbiosis or process of transplantation from one mind into another mind, than leaving it to nurture just in the place where they sprang up'.

V. Originality

According to the definition by Rastogi R.N. (1996), originality is the basic acumen of human being to produce, unusual, unconventional, novel, atypical and audacious answers to questions, least expected responses to problems, and unique unorthodox refreshing interpretation of issues, situations and events. Originality entails a person to avoid following the beaten track and always blaze a completely new trail. Even if the path is completely un-trodden, unpaved and difficult to negotiate and traverse, the quality of originality would entail one to choose a path less walked. A serendipitous mind is the most significant aspect of the person who is always original. The person with characteristics of originality would often discount the common approach followed by most of us as just one of the many possible approaches to the problem and never consider it as the best solution. While original thinking is to a great extent, is an innate quality of a person, with constant training and effort, it is always possible that one can develop this distinct quality with considerable effort and perseverance.

VI. Complexity Orientation

Going by the definition of Rastogi R.N.(1996), complexity orientation is the distinct characteristic of the person to confront, challenge and find meaning , in complex and ambiguous situations, and to enjoy the very act of analyzing, integrating, clarifying and resolving them. As a matter of fact, this is one of the essential attributes of any leader. A leader is one who can find simple solution for complex problems. Order out of chaos or reducing complexity to comprehensible simple things is indeed the way one would prefer things than letting simple thing go complex or letting orderly things eventually move into chaos. According to Machiavelli, the opportunities offered by a good crisis are enormous and very valuable, and one should never let go the opportunity thrown up. The leader recognizes the fact that the higher the complexity of the situation, more pressing is the need for immediate and qualitative solution with the possibility of a solution stretching the human potential to its limit. On the other hand, a simple situation with a simple problem can be routinely solved and it cannot be exciting to people who look for creative adventures. Complexity orientation is to a great extent contingent to the level of one's intellectual advancement and growth in emotional maturity. Whereas, the lesser beings or mortals are comfortable with simple situations, and would prefer to countenance simple problems, and are never at ease with

situations making them leave their comfort zones, and often balk at complex situations, a person of complexity orientation would exactly love the opposite situation. But for this category of people, who constantly look for challenges and complexities in life, the society and social structure would be stagnant and there would not have been improvement and progress at all.

VII. NEED FOR INNOVATION

The need for innovation arises from the fact that most of the human beings, i.e. even those working in business organizations which are well structured and societal organization which may not be so well structured, are overwhelmed by activities which are purely transactional in nature, and often they do not realize this fact. The everyday routine transactional works are so pressing and so large in volume, so much so, that an average individual thinks and sincerely believes that-- that is real work-- whole of work and nothing but this type of slogging work is that enables organization sustain and even progress. The truth cannot be farther from this notion. There could be occasions when one faces some kind of challenges initially in work, and this may be often, purely transactional in nature, meaning thereby a job requiring a bit of tweaking of the existing arrangement. And the employee gets over this challenging situation mostly by following the suggestions and the advice of one's peers. On a relatively fewer occasions, when an employee comes up with an innovative solution for the problems, for a while, the employee thinks that it is unusual response and momentarily lauds himself. Thereafter, the employee tends to rest on his laurel rather than continue with the quest for innovation for its own sake. In a situation where the external factors are quite stable and conducive, the employees, for most part of their life, can continue with this state of existence and lead a contented life. Only when the problems are to be solved from the perspective of fundamental change, arising out of complex dynamics coming into play, or disturbance or even disruptions in external conditions, there is a need for total transformation of the process, large scope for innovation suddenly appears on the horizon. With the boundaries of knowledge and technology are ever widening and expanding, even to remain in the same place in the competitive world one needs to innovate. The world does not change due to people who ordinarily comply without questioning, and prefer to go by precedence without thinking originally. When a person refuses to go by the beaten track and questions the existing procedures critically, people around him sit up and take notice of the person and are forced to examine the situations. And these are the situations that can cause transformational changes and, only transformational attitude can be the driving force for innovative thinking and progress. If the need for innovation is indeed an undisputed and unvarnished truth, the act of innovation is fully dependent on transformational thinking and action. Invariably the transformational thinking would imply, pushing the outer envelope of one's endeavor or domain of activity, farther and farther. And without innovation, one is doomed to stagnate, and ultimately would be wiped out entirely from the scene. And hence for progress and moving ahead, whether from the perspective of society or business, innovation is an imperative and a desideratum.

VIII. KINDS OF INNOVATION

Based on the value addition, Elaine Dundon (2007) famously categorized innovations into three types, namely, efficiency innovation, evolutionary innovation and revolutionary innovation. It must be noted that the division of innovation into three groups can never be presumed to be watertight and compact. Rather, it can only be said that the categorization based on value addition of innovation is a kind of continuum on the innovation value addition scale. Those innovations contributing in a significant way to the present state of the art technology is rightly termed as revolutionary --as it is a kind of drastic departure from the present line of thinking— which disturbs and even dislocates the existing state of affairs. Evolutionary innovation, on the other hand, is a process of building and leveraging on the present, without breaking away completely from the past. In a way, it is an approach of incremental work on evolutionary basis and the whole process is aimed at contributing moderate amount of value addition. Efficiency revolution on the other is purely incremental in nature and is aimed at cutting cost, improving efficiency, improving quality and providing a better appeal to customers and so on. As the effort is to achieve better efficiency and get more mileage from the existing situation, the whole approach is one of working around the present technology. It is therefore obvious that, both efficiency and evolutionary innovation operate, by and large, within the existing frame work of innovation without total break or complete rupture from the existing state and appears to be a kind of seamless progress along a continuous line. An entrepreneur, therefore, ideally would be required to straddle all the three types of innovation, moving from one mode to another seamlessly with ease according to the need and exigency of the situation. This particular attribute would help a great deal the entrepreneur in carrying out his plan of growth in the short, medium and long time horizons and take the venture too far.

IX. FOSTERING CULTURE OF INNOVATION

In view of the significance and importance of innovation, at all levels of entrepreneurship, business, society and government, it is necessary to assiduously and carefully cultivate, nurture and develop innovation both as a trait, credo and way of life. It must be remembered that innovation is not just an inborn quality but can also be nurtured and developed through systematic training by following established and proven techniques. There has been considerable work in this area by Edward De Bono(1993). While the exact techniques can be gleaned or studies from the literature, the underlying feature of all these techniques rests on the fact that one need to employ the full potential of divergent thinking-- apart from the conventional convergent thinking. The convergent thinking has always been the focus of development at the

childhood, schooling, secondary and tertiary education under the present system in majority of the cases. The process of convergent thinking revolves around clearly identifiable steps of problem identification and subsequent definition, diagnosis of the situation and the problem, and systematic analysis followed by prognosis, and synthesis providing solution to the problem. On the other hand, divergent thinking generally follows the process of identification of the problem on a tentative basis, searching solution starting with most familiar and then moving to complex ones, shifting one's perspective and frame work quite widely, all the time. To put it simply while convergent thinking is movement along closing spiral, the divergent thinking is movement along ever widening spiral. The search for solution is accomplished through a process of divergent movement along the locus of expanding concentric circles, moving from the simple obvious looking solution at the centre to more non apparent, novel and innovative solutions at the periphery. In the process, one might use analogies from other realms of study or knowledge. Indeed the ability to bring into play large number of diversity or variety or perspectives, in terms of ideas, imagination and creative outcome can be of great use. The right hemisphere of the human brain is associated or credited with intuitive, imaginative and creative thinking process and traditionally, in our quest for imparting reasoning, logical thinking-- which are all part of convergent thinking,-- not much attention is paid to the development of imagination and creative endeavor in childhood. It is therefore imperative, that one concentrates on development of this aspect right from the childhood by engaging in stimulating creative activities like singing, painting and story telling, even as one concentrates on inductive and deductive logical thinking aspects of one's development. The ability to entertain and examine all ideas without necessarily accepting them is one of the important requirements for divergent thinking. As Mahatma Gandhi said, "I do not want my house to be walked in on all sides and my windows to be stuffed. I want the cultures of all lands to be blown about my house as freely as possible. But I refuse to be blown off my feet by any."

X. The Importance Of Innovation For Contemporary Life

The astonishing progress of mankind in this world has been mainly due to ability of the human being to think and innovate constantly. The locus and trajectory of the path of progress is solely determined by the quantum and quality of springs of innovative-thinking in a society, organization or establishment. We have witnessed societies and organizations prospering when the emphasis and stress is on promotion of innovative thinking and innovation at every level. On the other hand, we have equally large number of examples, of organizations and societies languishing, remaining undeveloped or even decaying when they stopped innovating. The process of innovation is not necessarily confined to technology, arts, sports, literature, science, etc. and covers every aspect of human endeavor. Indeed the boundaries of innovation are never fixed and keep expanding with time. Innovation is ever welcome and is essential in every sphere of the individual growth namely, physical, sensory, emotional, mental, intellectual and ultimately even in spiritual. Innovation in every sphere, therefore, is the process of human being seeking and attaining a higher level of consciousness than hitherto achieved and this process is never complete. For the true measure of human potential is still unknown, or at best known very little and every innovation is thus a step towards knowing our own true self and its depth.

While the importance of creativity and innovation in the field of technology and business seems very obvious and is readily accepted as the key determinant for success at individual and organizational level, what is not so obvious is, the significance and importance of innovation in every walk of life or avocation and its all pervasive and ubiquitous nature. Innovation, thus, should become an ingrained character of an individual. The life would then become a matter of series of interesting challenges to be overcome with interest and passion by the individual. The joy and fulfillment in identifying and confronting a problem and solving to the satisfaction of everyone, is indeed a true inner joy, which only human being with higher level of consciousness can indulge in and appreciate. Thus, , being totally absorbed and indulging in exercises in creativity and innovation for its own sake, is rewarding to human kind. For innovation implies constant learning and, without learning society stagnates and putrefies. As John Naisbitt has remarked, "In a world that is constantly changing, there is no one subject or set of subjects that will serve you for the foreseeable future, let alone for the rest of your life. The most important skill to acquire now is learning how to learn." And that indeed is the key to inculcate innovative spirit in individual, organization and society.

XI. Conclusions

Technology is the cutting edge of growth and progress of society. The advancement of society to a large extent, hinges on the ability of the society to leverage its technology and intellectual capital. If one looks at the pace at which technology is growing, it is, at once, evident that the innovation is the engine which propels continuously the advancement of technology along its high speed trajectory. A teacher who does not grasp the significance of innovation in technological advancement and its role in engineering education is obviously at a great disadvantage. Just as technology calls for constant innovation, the teaching of technology or engineering education calls for sustained and constant innovation. Indeed viewed from this angle, the innovation in engineering education is more challenging than innovation in technology. What else can be a better place to learn innovating in engineering, other than the engineering college where the student cuts his teeth and to which students come with great hope? The culture of innovation should therefore rightly start at colleges with engineering education taking the lead and one need to pay greater attention to this important aspect.

REFERENCES

[1] De Bono, E,(1993) Parallel Thinking, Viking London

[2] De Bono, E, Sur/petitions, Harper Collins,New Delhi 1993

[3] Elain Dundon,(2007) The Seeds of Innovation—Cultivating the Synergy that Fosters New Ideas. Prentice Hall of India Private., Ltd., New Delhi