Estd :2005
MALLA REDDY COLLEGE OF ENGINEERING
(Formerly CM Engineering College)
Approved by AICTE - New Delhi, Affiliated to JNTU - Hyderabad,Accredited by NBA & Accredited by NAAC.
ISO 9001:2015 Certified Institution, Recognition of College under Section 2(f) & 12 (B) of the UGC Act,1956.

INTERNATIONAL CONFERENCE ON TRENDS IN INFORMATION, MANAGEMENT, ENGINEERING AND SCIENCES (ICTIMES-2019)

DATE :DEC 20th 2019 and DEC 21st 2019

Malla Reddy College of Engineering Secunderabad

Principal
Malla Reddy College of Engineering
PRINCIPAL

ICTIMES 2019

5th INTERNATIONAL CONFERENCE ON

Trends in Information, Management, Engineering and Sciences

Proceedings of International Conference on
Emerging Technologies in Computer Science (ICETCS)
- ISBN: 978-93-88808-61-3

Convenor
Dr. M. Narayanan
HOD-Dept. of CSE

Editor
Dr. J. Gladson Maria Britto
Professor - CSE

Editor in Chief
Dr. P. John Paul
Principal

Organized by

Estd :2005

MALLA REDDY COLLEGE OF ENGINEERING

Approved by AICTE - New Delhi, Affiliated to JNTUH, Accredited by NBA & Accredited by NAAC.
ISO 9001:2015 Certified Institution, Recognition of College under Section 2(f) & 12 (B) of the UGC Act,1956.

website: www.mrce.in

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

SMART WAY TO COMMUNICATE THE PASSENGERS IN THE TRAIN USING THE HYBRID TECHNOLOGY

Dr.T.Sunil
Sunil.tekale2010@gmail.com

U. Nagaiah
nagaiah1212@gmail.com

Abstract:

In order to cater to the needs of passengers travelling in the train by reserving the seats, the system is designed so as to provide passenger specific information as well as common information to all the passengers travelling. Passenger specific information can be like information in advance about the destination along with the information of the live train. Whereas the common information to the passengers can be like information about the live train. The system is designed such that it works on time every time because of the technology used which is called the hybrid technology. Here combination of internet and intranet technology is used. The passengers can avail this facility by opting the same during the seat reservation process.

Introduction:

Providing information to the passengers in the train is very important now days because of various reasons. Which is done and the facility is available to the passenger with the help of internet. But I want to provide the facility to the passenger who is travelling without making use of internet directly. The touch screen device will provide information related to location of the train and will also provide information about the destination of the passenger. The touch screen device when touched will provide this information. Apart from that it will also alert to the passenger about the destination in advance by means of alarm. As internet connectivity may not be available at all location especially in the forest area, this method of using intranet communication will provide the best solution for the same. This is an innovative method of providing information /alertness to the passenger without using the mobile device.

The basic information of passengers performing reservation will be available with the railway authority and this has to be exported to the main server arranged for a particular train. The server has to import the data and has to transfer the same to various clients which are located at every passenger seat along with the common client for each block of the bogie. The purpose of client common to block of each bogie can be referred as common client system is to provide general information common to all the passengers , whereas the purpose of the individual client is to provide particular information to specific client . The server will be updated every time from the main server and the same information will be used to provide the alerts and alarm to the specific passenger. The communication between the server in the train and the client can be done using intranet communication system and between two servers with internet. The basic idea which is novel is to see that the information which has to be provided to the passenger should be given on time and every time. The various devices used are touch screen devices with alarm facility (speakers) referred as clients along with the hubs at every bogie and the main server in the train. The communication can be with the help of

ISBN 978-93-88808-61-3

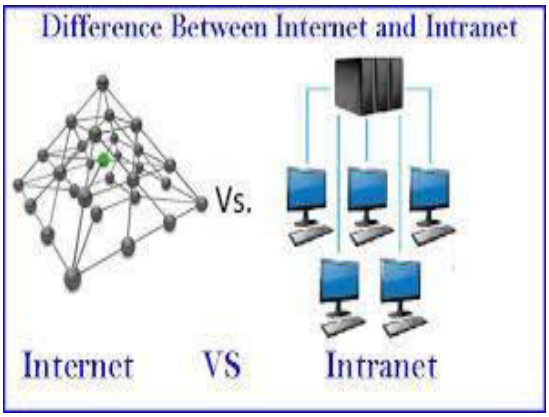
wires as well wireless. So as soon as the train is ready for departure at source station the information will be imported and the same can happen at multiple other stations. The passenger will be given an option to select the facility while the passenger performs the reservation. This facility will be provide only to the passengers travelling by purchasing tickets in advance (reserved). Where as in general compartments the arrangement can be done only for the common clients which will inform about the position of the live train.

Design consideration:

1. The system is designed with the help of internet and intranet technology to provide passenger specific and common information.
   a. As the chances of designing with internet technology may fail in many locations.
   b. The system is designed with touch screen devices so as to make it easier to use for the common man or for a passenger who is not well versed with technology.
   c. The system will provide different information to the clients according to the destination chosen.
   d. The System designed helps for specific individual as well as common to all.
   e. Usage of hybrid technology is novel so as to see that the system works on time every time.
   f. The system is easily managed because of having total control of data.
2. The system is designed using intranet so that the same works in all the geographical locations where internet do not /may not work.
3. A system as said in claim 1 is designed by using touch screen device is an innovative idea/novel idea as the common man faces problems with the latest technology.
4. The system designed is very innovative in terms of providing information as the destination chosen by the passenger. The system will alert and guide the passenger in advance, so that there will not a problem at the last minute.
5. The designer of the system is very complicated and innovative as the same system has to provide information to individual passengers when prompted as well as to all passengers sitting in the compartment at specific time interval.
6. Designing the system with hybrid technology was innovative concept so that we can have better and concrete system at place.
7. The algorithm used here will help the build a robust system.
8. The system is designed so as to cater to the need of people who are not well versed with technology.

Fig.1 Image of Intranet used in the process of passenger communication

Fig.2 Image of internet and intranet used in the process of passenger communication



Challenges:

To store the data from the source station before the train starts and to feed the same in the main server which will be placed in the drivers cabin and to see that the cables laid out for passenger communication are not disturbed or cut by any means. The touch screen device should be made very tough or robust so that it will be damaged easily by the passenger wanted or by mistake.

Existing System:

In the existing system the passengers are made to depend on the mobile phones or they need to depend on another passenger to provide information regarding the current location of the train as well as to wake up in the morning for the destination station. The problem here is if the mobile network is not working than there is every chance of system being collapsed. So in order to reduce the dependency the new system will help to provide information to the passenger as and when they required as well as they can get the information half an hour before they actually have to get down the station.

How the system works:

As the system uses intranet and internet technology which otherwise can be called as hybrid technology it is very easy to provide information to the passenger in much better way. The information will be collected at the source station with the help of internet facility and the information will be shared among the various passengers with the help of intranet facility along with the touch screen devices arranged.

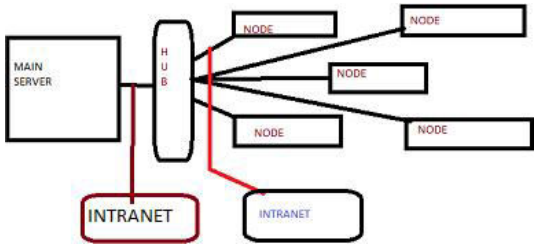
The information which is the data pertaining to the passengers reserved for that particular train will be gather and stored in the server system at the source station, then all the nodes which are connected to the hubs in various bogies are connected. The nodes will get the information from the server using intranet technology. So here the data is stored in the server with the help of internet and the same is shared among various nodes using intranet technology, then the dependency on the mobile technology and mobile network is completely zero.

Fig: 3 shows the touch screen which will be used by the passenger.

Results:

The data is stored in the server system from the main railway reservation system using internet technology which can be referred as MD1 and then using the intranet technology between the server and the various hubs and the nodes connected to the hubs the data is transferred from the main server to the nodes via hubs which can be referred as ND1…NDn.

Fig: 4 Refers to communication between the main server and the nodes present in various bogies.



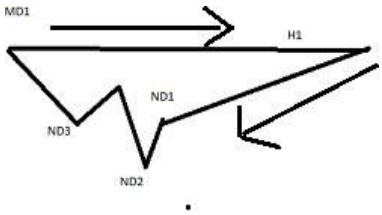
MD1-> ND1 MD1->ND3

MD1->ND2 MD1->ND4

So the information is passed from the main server i.e MD1 to rest of the nodes including the hubs and accordingly the passenger can get the information regarding the various stations and the distance from the destination.

MD1>H1>ND1 MD1->H1->ND2

Fig:5 How the data flows from MD1 to H1 and from there ND1…NDn



Information between the MD1 and the H1 which is the hub is communicated using intranet and from there to the various nodes is also by intranet technology only.

Whereas the information from the main reservation server to the main server in the train is communicated using internet technology.

Conclusion:

When the system is designed it will be very useful for the passengers who are travelling by taking the upper part of the berth because there is no other way for them to get the information related to the places they are visiting and arriving. This particular system will help them to get information regarding the position of the train and also can get the information about their destination in advance. The passenger for this may have to shell minimum amount which can be used for the development of the system in future. Overall the conclusion is that this is going to be well defined asset for the passengers as well as the railway board as it helps to increase the revenue and passengers will be happy because o getting right information at right time. The passengers will be provided with touch screen for this purpose.

REFERENCES

[1] Carol L. Schweiger, "Real-Time Bus Arrival Information Systems - A Synthesis of Transit Practice",

Transportation Research Board, 2003.

[2] "Review of Current Passenger Information Systems", Prepared for the INFOPOLIS 2 Project (No. TR 4016),
Deliverable 1, WP03, Info polis 2
Consortium, August 1998.

[3] Hu, K. and C.K. Wong, "Deploying Real- Time Bus Arrival Information and Transit Management Systems in Los Angeles", abstract prepared for the ITS America
12th Annual Meeting, Long Beach, Calif., April 29-May2, 2002.

[4] Helsinki City Transport System http://www.hel2.fi/ksv/entire/repPassenge rInformation.htm

[5] Tolarno Inc. - Passenger Information Services http://www.telargo.com/solutions/passene r_informationservices.aspx

[6] Terron Microsystems Pvt. Ltd.- GPS based Passenger Information System for Buses http://www.terronmicrosystems.com/prod ucts.php

[7] Brendan Kidwell, "Predicting Transit Vehicle Arrival Times", GeoGraphics Laboratory, Bridgewater State College, August 2001.

[8] Lin, W.H. and Zeng, J., "Experimental Study of Real-Time Bus Arrival Time Prediction with GPS Data."

Transportation Research Record, 1666, pp. 101-109, 1999.

[9] Chien, I. J. S. and Ding Y., "Applications of Artificial Neural Networks in Prediction of Transit Arrival Times", 1999 Annual Meeting of ITS America, Washington D. C., 1999.

[10] Wanli Min et al, "Road Traffic Prediction with Spatio-Temporal Correlations", IBM Watson Research Center, 2007.

[11]A. Guin, "Travel Time Prediction Using a Seasonal Autoregressive Integrated Moving Average Time

Series Model", IEEE ITSC, 2006.

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

An Efficient method to Transfer Files in Peer- to-Peer Networks over Video on Demand Services

[1]M.Narayanan, [2]E.Lingappa
[1]Professor, [2]Assistant Professor
Department of Computer Science and Engineering,
Malla Reddy College of Engineering, Maisammaguda,
Dhulapally, post via Kompally, Secunderabad, Telangana

Abstract

One of the most promising technology is Peer-to-Peer. Each peer acts as a client and server so every peer is accounted for the downloading and uploading of files. Though frequently compared to client server model it is altogether distinct to it. Therefore it is also known as single system program so at that instance each peer can act as client/server. Content distribution or file sharing or dissemination of information of the file is one of the most important application. Besides this, the other applications include publications and dissemination of software, content delivery network, streaming media and multicast streaming. It also allows on demand content delivery. Science networking, searching and communication are other applications of peer to peer.

Keywords: P2P Network, Client server Model, Hybrid Model, Searching Communication

I. INTRODUCTION

The concept of Peer-to-Peer file sharing technology is growing at an unprecedented rate. Each peer can perform like both client and server and that sets them apart from the client server model. File can be shared depending upon the bandwidth and the peer possessing larger bandwidth is given the precedence followed by the other. Suppose the key server is sharing the file to client and another peer requests for file sharing, the key server will put together an arrangement for the requested peer. And the client receiving the file at this time will transfer the file to the client which is making the request. In this scenario it is free for anyone to operate as a client and server. Therefore the amount of the file transferred will be larger and at a higher speed. It comes handy for downloading a video file from the net. The files can be shared by utilizing the techniques such as MCC, FIFO, LFU, and LRU. When a file wishes to share from the server, it will apparently look for the peer which possess a higher bandwidth [15].

Peers with similar bandwidth experiences LFU (low frequently used). So the peer which is not used frequently will receive the file. In peer to peer the files are stored in the cache memory in the server. When the cache memory is completely occupied it will then erase the files on the basis of first in first out (FIFO). The file which is deleted first will be the one which was first in. It is the responsibility of the cache to locate the client having the necessary file. This memory enables the files to be transferred between peer to peer. The capacity of the cache memory is up to 1024mb equivalent to 1kb.It will follow the following technique when the cache is completely occupied. Peer to peer enables many users to work simultaneously. Therefore it is considered as the best file sharing network. Depending on the bandwidth, the peers will be separated as clusters [2] with each cluster consisting a definite bandwidth.

For example cluster1 contains 1mbps bandwidth, cluster 2 contains 2mbps and similarly each clusters comprises of different bandwidth. Within each cluster many peers will be separated wherein every peer will possess similar bandwidth. In this peer to peer, all the peers will be linked in corresponding manner [1], [8]. There is no restriction as who should act as a client and who should be a server. This property enables to not only share video file but even some digital files and some computer files, books, movies, music and games. This particular P2P software enables to look for client at nearest proximity in the P2P network and conveys the file. The Fig.1 shows the client Server Model and peers (nodes) of P2P network will be end user computer system interconnected between the internets. Scientist Gnutella and kazaa were the first to develop this new P2P file sharing method.

ISBN 978-93-88808-61-3

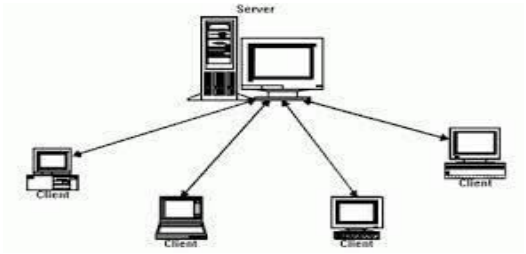


Fig 1: Client Server Model

II. RELATED WORK

In the year 1999, the earlier peer to peer file sharing method was developed by NAPSTER. Following the death of NAPSTER, gnutella and kazaa developed the new peer to peer file sharing network [7] . The old Peer to Peer file sharing was based on the central server system and enabled sharing of only music files effectively. Following extensive research, the system has developed to a great extent and any kind of files can be transferred utilizing this peer to peer file sharing technique. Unlike the client server model, this method is connected parallel and is known as decentralized and dispersed design [3][1].

A. Mentioned below are the two types in P2P

The structured p2p is organized by some definite standard and some algorithms. Some of the real time examples of structured P2P are some of the network computer workstations [5]. Unstructured P2P computer architecture consists of three types of models as follows: It is a democratization of every peer group nodes. They are two forms in order to accomplish the route:

One of the probable structuring is direct messaging. It is communicated through peer group members until an object of the member group is found. Further it must establish the members in HORIZON group. Horizon means limit of visibility from the node generating the query.

Another probable way for attaining the routing structure is distributed catalog. It necessities an energetically balanced catalog because it is indexed as parameter and searches as distributed catalog. This may also be not very efficient but it is very safe to work.

It is very necesary to enhance the P2P models in order to improve the search potential and system performance. Data accumulation in Peer-to-Peer is enormous. So in order to address these issues, this paper presents the application of data mining technology to P2P network. Based on SWLDRM, few developments are considered and nodes are clustered depending on the characters of object stored by K-NN algorithm. Simulation result proves that K-chord is more efficient in terms of performance and search when compared to SWLDRM [9].

One of the extensive classification of community based P2P systems is presented in this paper. Users belonging to a definite network forms this community. And the augmentation of these communities is influenced by factors such as value of the content, projection for enhanced performance and user experience enhancement. A campus network and a national ISP located in diverse continents are the two distinct environments that this study focuses. Here, the key P2P systems are found to be large scale closed communities. Results confirm that traffic on Internet peering links are reduced by localizing traffic inside ISP boundaries [10].

It is becoming increasingly challenging for ISPs because of the unprecedented growth of P2P applications consuming humongous bandwidth resources. So, it is paramount to handle P2P traffic efficiently and simultaneously protecting the P2P user interest. This paper primarily focus on analysing current mainstream P2P optimization strategies and implement NRDA technology. This technology permits ISPs to manage P2P traffic efficiently and independently on certain links and also to respond quickly in case of an abrupt necessity of resources. By and large, the planning of network resources are executed by the NRDA and it also offers resource planning capability to the ISPs [11].

In today's internet world, ISPs are facing an uphill task of providing basic network services for P2P users and also to effectively manage network bandwidth usage. But, existing strategies fail to fulfil these requirements. This paper proposes to devise a plain and efficient system for ISPs to strike a balance between service and network management. This can be attained by suggesting a file-aware P2Ptraffic classification method which can identify files and the associated flows. Two alternatives are proposed. One is by limiting the per-file bandwidth consumption and the other by measuring a real-life trace from peers and files perspective. The results show that as per the actual demands, ISPs can expediently choose suitable traffic management parameters [12].

Most of the video streaming services are developed for wired networks. But the challenge is to stream it in wireless environment which demands several alterations. Through an logical representation, this paper proposes a performance evaluation model of the traffic behaviour which bears a resemblance to the network interactions during a video transmission. In wireless environment, when the number of nodes increases the quality of video degrades due to collisions. Therefore, *P2P-TV* applications should be integrated from lower layers to meet the level of quality requirements. The throughput parameter is determined by this model and the network performance is evaluated precisely [13].

There should be monitoring of the Service Level Objectives in order to meet the Service Level

ISBN 978-93-88808-61-3

Agreements (SLAs) and regulate vital network services. However, the probing techniques are expensive and are also labour-intensive and prone to error. So, Peer-to-Peer (P2P) technology is employed to improve the detection of SLA breach. A P2P management overlay is considered to coordinate the probe activation and share measurement results between the network devices. In large scale networks, an autonomic P2P solution is proposed to coordinate active measurement probes. The solution is proved to be feasible as per the simulation results [14].

B. Centralized P2P

It consists of a central look up server linked in a star network style. In this type of peer to peer model, the message can be sent with ease and with greater speed because of less traffic. Because of the fact that it has less traffic it can be quickly addressed. But bottleneck behaviour which is a single point of failure is the biggest drawback. Though addressing is very efficient it is not safe to work. The Fig.2 shows the centralized P2P [4].

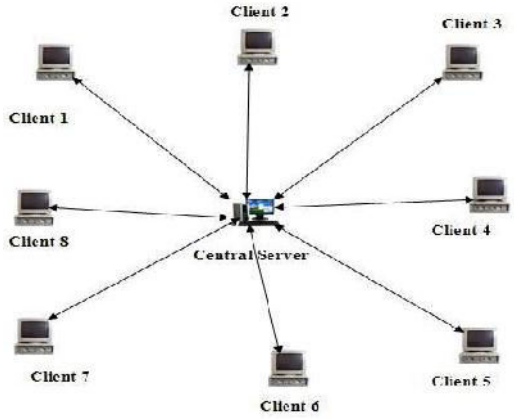


Fig.2: Centralized P2P

C. Hybrid P2P

It is an overlay routing structure consisting of super peers and leaf nodes. Hence leaf nodes are connected in star network and super peers act as shield to the leaf nodes [8]. Addressing is partly efficient and it is not safe.

Furnished below are few of the benefits of the peer to peer network:

a. Operating system is not needed in this network.
b. Unlike the client server model, a specialized complex network set up is not required. It can be created with ease and does not demand any superior knowledge.
c. Server is not so pricey because of the fact that an individual client can act as a server when uploading a file to another user.
d. If one of the peers does not succeed in sharing, it will not disturb any other part of the network. It would be just occupied with another work of sharing and will remain unavailable to other peers.
e. Compared to the client server model, the file sharing will be at a higher speed.
f. Easy availability and reduced cost is another big benefit of peer to peer network. So it enables users to utilize it in low cost.
g. Peer to peer has well tested peers, it will not have the peers which is not ready to share the file among another peer i.e. well tested simplicity.
h. It doesn't require a dedicated server. Anyone can act as server and anyone can act as client, so any computer can access both server and workstation.

D. Problems of P2P networking

Some of the potential problems faced by the users with Peer to Peer software are Bandwidth utilization, copyright infringement, and security issues. It also face some troubleshooting network problem, A computer may fail to process for many reasons this one of the basic reason for the problems in P2p networking. If it is not working problem it will affect the entire home networking to stop functioning. The Fig.3 illustrated the model diagram of P2P Networks.

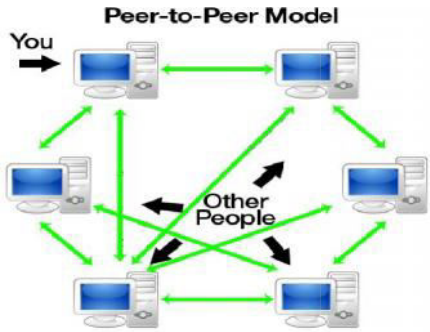


Fig.3: Model diagram of P2P networks

III. PROPOSED MODEL FOR P2P SYSTEMS

A. Core Transfer Engine Layer

The requested files between peers is transferred by this layer and it also carry out all the responsibilities of Peer actions. In this particular project, we anticipate to see some codes around Peers. This part is the heart of this system. When a peer commences its work, the first task is to register itself as a peer. Subsequently it should play the part of both server and client [6] Later, if any peer asks a file, first it should search the file and after receiving the file's information such as the destination peer host name, it should utilise that information to connect to the peers and then download the file. The PNRP Manager (Peer Name Resolution Protocol) class is responsible to Register and Resolve peers.

ISBN 978-93-88808-61-3

The Register () method registers the peer in the cloud and accepts a list of Peer Info type as its input argument.

B. Proposed Algorithm: Algorithm for Peer Registration

Algorithm_PeerInfo_Registration
Begin
List PeerInfo, RegisteredPeer;
ForEach Registration;
String TimeStamp;
String TimeStamp = String.Format("FreeFile Peer Created at :Zeroth Position") DateTime.Now.ToShortTimeString ();
Registration.Comment = TimeStamp;
Try
Registration.Start();
Begin
IF RegisterdPeer.FirstOrDefault
X is HostName is
Equal to Registration_PeerName
And PeerHostNamePeerInfoPeerInfo =
New PeerInfo (registration.PeerName.PeerHostName,
Registration.PeerName.Classifier,
Registration.Port);
PeerInfo.Comment equal to Registration.Comment;
RegisterdPeer.Add(peerInfo);
EndIF
End
EndFor
End

C. File Transferring service in P2P systems

In order to give the essential files to other peers, the File Transfer Service Host class enables each peer as a server host. The TCP protocol is utilized for conveying the data among peers. Depending on the peer host name the DoHost() method gets an address. Subsequently an interface is added who applied the Service Contract feature. Hence, to make its methods reachable around service each peer publishes a service to the outside world. The Table 1 illustrated algorithm for file transferring service in P2P systems

TABLE 1: ALGORITHM FOR FILE TRANSFERRING SERVICE IN P2P SYSTEMS

```
Sealed class FileTransferServiceHost
{
   Public void DoHost(List<PeerInfo> peers)
   {
      Uri [] Uris = new Uri[peers.Count];

      String Address = string.Empty;
      For (inti = 0; i<peers.Count; i++)
      {
         Address = string.Format("net.tcp://{0}:{1
}/TransferEngine",
peers[i].HostName, peers[i].Port);
         Uris[i] = new Uri(Address);
      }

FileTransferServiceClass
currentPeerServiceProxy = new
FileTransferServiceClass();
ServiceHost _serviceHost = new
ServiceHost(currentPeerServiceProxy, Uris);
NetTcpBindingtcpBinding = new
          NetTcpBinding(SecurityMode.None);     }
}
interfaceIFileTransferService
{
   [OperationContractAttribute(IsOneWay = false)]
byte[] TransferFileByHash(string fileName,string
hash, long partNumber);

   [OperationContractAttribute(IsOneWay = false)]
byte[] TransferFile(string fileName, long
partNumber);
}
```

IV CONCLUSION

One of the most promising technology in the internet world is the P2P network. After the multi-core processor is developed the P2P network will achieve unprecedented worth in the web. This study is a complete survey paper providing details about P2P networks. My survey provides types of P2P computing algorithms. Currently every field depends on computer and its various application. Therefore it is highly essential to design state-of-art P2P systems for video on demand service.

REFERENCES

[1]. Biazzini, M.; INRIA-Bretagne Atlantique, France; Serrano-Alvarado, P.; Carvajal-Gomez, R.,"Towards improving user satisfaction in decentralized P2P networks", in Collaborative Computing: Networking, Applications and Worksharing (Collaboratecom), 2013 9th International Conference Conference , 20-23 Oct. 2013.

[2]. Garant, D. ; Dept. of Comput. Sci., USNH, Keene, NH, USA; Wei Lu,"Mining Botnet Behaviors on the Large-Scale Web Application Community ", in Advanced Information Networking and Applications Workshops (WAINA), 2013 27th International Conference ,25-28 March 2013.

[3]. Hirave, T.; Dept. of Comput. Eng., Mumbai Univ., Mumbai, India ; Surve, S. ; Malgaonkar, S.,"Selecting efficient peers in P2P networks for parallel task computing",in Advances in Technology and Engineering (ICATE), 2013 International Conference ,23-25 Jan. 2013.

[4]. Biazzini, M. ; INRIA-Bretagne Atlantique, France ; Serrano-Alvarado, P. ; Carvajal-Gomez, R.,"Towards improving user satisfaction in decentralized P2P networks", in Collaborative Computing: Networking, Applications and Worksharing (Collaboratecom), 2013 9th International Conference , 20-23 Oct. 2013.

[5]. Garant, D. ; Dept. of Comput. Sci., USNH, Keene, NH, USA ; Wei Lu,"Mining Botnet Behaviors on the Large-Scale Web Application Community ", in Advanced Information Networking and Applications Workshops (WAINA), 2013 27th International Conference ,25-28 March 2013.

[6]. Hirave, T. ; Dept. of Comput. Eng., Mumbai Univ., Mumbai, India ; Surve, S. ; Malgaonkar, S.,"Selecting efficient peers in P2P networks for parallel task computing",in Advances in Technology and Engineering (ICATE), 2013 International Conference ,23-25 Jan. 2013.

[7]. Ripeanu, M. ; Dept. of Comput. Sci., Chicago Univ., IL, USA,,"Peer-to-peer architecture case study: Gnutella network", in Peer-to-Peer Computing, 2001. Proceedings. First International Conference , 27 Aug 2001-29 Aug 2001.

[8]. Kang Chen ; Dept. of Electr. &Comput. Eng., Clemson Univ., Clemson, SC, USA ; HaiyingShen ; Haibo Zhang, "Leveraging Social Networks for P2P Content-Based File Sharing in Disconnected MANETs" , Mobile Computing, IEEE Transactions on (Volume:13 , Issue: 2), Date of Publication : 26 November 2012

[9]. Zhu, Xiaoshu, Miao Xie, and Tang Lu. "Application Research of Data Mining Techniques in P2P Network." In *Genetic and Evolutionary Computing, 2009. WGEC'09. 3rd International Conference on*, pp. 297-300. IEEE, 2009.

[10]. Torres, Ruben, Marco Mellia, Maurizio M. Munafo, and Sanjay G. Rao. "Characterization of community based-P2P systems and implications for traffic localization." *Peer-to-Peer Networking and Applications* 6, no. 2 (2013): 118-133.

[11]. Ma, Dongchao, Xiaoliang Wang, Wenlong Chen, Shen Yang, and Li Ma. "A research on dynamic allocation of network resources based on P2P traffic planning." *Peer-to-Peer Networking and Applications* 7, no. 4 (2014): 511-524.

[12]. Song, Tian, and Zhou Zhou. "File-aware P2P traffic classification: An aid to network management." *Peer-to-Peer Networking and Applications* 6, no. 3 (2013): 325-339.

[13]. Urrea Duque, Juan Pablo, and Natalia Gaviria Gomez. "Throughput analysis of P2P video streaming on single-hop wireless networks." In *Communications (LATINCOM), 2014 IEEE Latin-America Conference on*, pp. 1-6. IEEE, 2014.

[14]. Nobre, Jéferson C., Lisandro Z. Granville, Alexander Clemm, and Alberto Gonzalez Prieto. "On the Use of Traffic Information to Improve the Coordinated P2P Detection of SLA Violations." In *Advanced Information Networking and Applications (AINA), 2014 IEEE 28th International Conference on*, pp. 613-620. IEEE, 2014.

[15]. M.Narayanan, C.Arun" An Efficient Method for Handling Data Segment with Multi-Level Caching over Video-on-Demand using P2P Computing", *European Journal of Scientific Research*. ISSN 1450-216X Vol. 93 No 2 December, 2012, pp.206-213

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

IOT BASED SMART AGRICULTURE MONITORING AND COMTROL SYSTEM

[1]Dr.J.GladsonMariaBritto,[2]Dr.Bhoopathy.V,
[1]Professor,CSE, Malla Reddy College of Engineering
[2]Professor,CSE, Malla Reddy College of Engineering

Abstract: The Major developing advancement technology in upcoming future is Internet of Things, commonly known as IOT is a promising area in technology that is growing day by day. Agriculture plays vital role in the development of agricultural country. In India about 80% of population depends up on farming and one third of the nation's capital comes from farming. The problems based on agriculture have been always hindering the development of the country. The highlighting features of this project includes wireless network sensors to connect multiple sensors data and to display big-data through Thing speak channel software to perform tasks like weeding, spraying, moisture sensing, bird and animal scaring, keeping vigilance. The development includes smart irrigation with smart control and intelligent decision making based on accurate real time field data. Which also includes temperature maintenance, humidity maintenance and weather reports. Controlling of all these operations will be through any smart mobile or computer connected to Internet and the operations will be performed by interfacing sensors. The data can be completely updated faster when compared to other wireless computing.

Keywords: *Internet of Things, Wireless sensor Networks, Micro keil version, Thing speak, Big data collection, cloud computing.*

INTRODUCTION

Agriculture is the unquestionably major process provider in India. With rising population, there is a need for increased agricultural production. In order to support greater production in farms, the requirement of the amount of fresh water used in irrigation also rises. Currently, agriculture accounts 93% of the total water consumption in India. Unplanned use of water continuously results in wastage of water. This suggests that there is an urgent need to develop systems that prevent water wastage without imposing pressure on farmers. Agriculture is considered as the basis of life for the human species as it is the main source of food grains and other raw materials. It plays vital role in the growth of country's economy. It also provides large ample employment opportunities to the people. Growth in agricultural sector is necessary for the development of economic condition of the country. Unfortunately, many farmers still use the traditional methods of farming which results in low yielding of crops and fruits. But wherever automation had been implemented and human beings had been replaced by automatic machineries, the yield has been improved. Hence there is need to implement modern science and technology in the agriculture sector for increasing the yield. Farmers could be able to smear the right amount of water at the right time by irrigation. Avoiding irrigation at the wrong time of day, reduce run off from overwatering saturated soils which will improve crop performance. The available traditional methods of irrigation are drip irrigation, ditch irrigation, sprinkler system. This problem can be easily rectified by making use of the automated system rather than the traditional systems. The current

irrigation methodology adopted employ uniform water distribution which is not optimal. The client and server programming condition which could improve the data over big data collection` Large number of entries could be over seen with online portal services. In addition to the standalone monitoring station, Wireless Sensor based monitoring system been developed which is composed of number of wireless sensor nodes and a gateway. This system here provides a unique, wireless and easy solution with better spatial and temporal resolutions. This could also include the farm security from animals attack without injuring the animals like in manual method.

Motion detector is used to sense the temperature of the animals and to be thrown away from farm land. The farmer is notified about the decision whether to irrigate or not through a either a web app or mobile app which is developed using WEB. Based on the decision received from the machine learning process, The farmer can trigger the irrigation process through his mobile phone .same is also provided through a web interface.

LITERATURE SURVEY

The older method and one of the oldest ways in agriculture is the manual method of checking the parameters. In this method the farmers only by themselves verify all the parameters and calculate the readings that's why to overcome this stress and relief from stress, It focuses on developing devices and tools to manage, display and alert the users using the advantages of a wireless sensor network method. It aims at making agriculture smart and modern using automation and IoT technologies. It provides a low cost and effective wireless sensor network technique to acquire the soil moisture and temperature from various location of farm and as per the need of crop controller to take the decision whether the irrigation is enabled or not. It proposes an idea about how automated irrigation system was developed to optimize water use for agricultural crops. In addition, a gateway unit handles sensor information. The atmospheric conditions are monitored and controlled online by using Ethernet IEEE 802.3.It is designed for IoT based monitoring system to analyze crop environment and the method to improve the efficiency of decision making by analyzing harvest statistics. The source of power can be powered by photovoltaic panels and can have a duplex communication link based on a cellular-Internet interface that allows data inspection and irrigation scheduling to be programmed through a web page .Various techniques agricultural applications like seed sowing, sloughing, water irrigation, crop cutting and etc. like this several operations were done with IOT. Various companies in INDIA and globally have been proposed in using micro controller based controllers for various have come with novel solutions using automated systems for various application with specific individually (www.smartagriculture.com). Most work carried out in literature and organizations have their inherent advantages and disadvantages. These manufacturers do not have multiple agricultural applications integrated in a single hardware. To eradicate such errors or disadvantages we are introducing a multi functional design using wireless sensor networks. The system was based upon an automated irrigation system by using mainly a soil moisture sensor and an Android smart phone. With this system, people can have a better control on their irrigation time and can also save water. In this prototype, different soil samples and crops for calibration at various moisture levels was tested. However, to improve this analysis, various soil samples from different places could have been tested and also during different weather conditions. Apart from soil moisture, other factors of the soil could have also been monitored. The cloud computing that could improve the advanced technologies using big data

ISBN 978-93-88808-61-3

collection through the fast updates of data's through online entries. In an updated wireless network sensor system, the data that could be updated through faster applications.

PROPOSED SYSTEM

The system is a combination of hardware and software components. The hardware part consists of programming microcontroller AT-mega328 which could connect the other sensors to collect the data. Solar panel which act as an renewable source of energy that is to be connected as an rechargeable power source to save the power. The moisture sensor that could recover the dry or wet condition of the soil and thereby the intimation could be sent to the farmers through web browser or GSM module. Then the centrifugal pump could be turned ON/OFF by the farmers from anywhere or anyplace through the online channel creation using Thing speak. Thing speak is the webpage designed using PHP. The webpage is hosted online and consists of a database in which readings from sensors are inserted using the hardware. soil moisture sensors along with LM38 comparator modules were placed in different soil conditions for analysis. It reads the moisture content around it. A current is passed across the electrodes through the soil and the resistance to is made the current in the soil determines the soil moisture. If the soil has more water resistance will be low and thus more current will pass through. An Intelligent IOT Based Automated Agriculture has been proposed so as to reduce the wastage of water and security to the crops. The system mainly monitors the behavior of soil moisture, air humidity, air temperature and secures the crops from animal attack and sees how it contributes to evaluate the needs of water in a plant. The data is taken from the sensors and is transferred through internet to the mobile application or web app and water pump turn ON/OFF using web app.

BLOCK DIAGRAM:

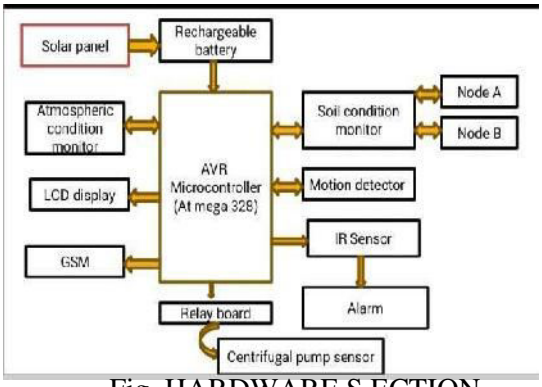


Fig. HARDWARE S ECTION

Which also includes temperature maintenance, humidity maintenance and weather reports. Controlling of all these operations will be through any smart phone will be performed by interfacing sensors. The WI-FI module that could be interconnected with the moisture sensor at various nodes of node A and node B. This collects the data continuously and to be entered through online portal using thing speak software. The major advantage of this method is secured and maintained complete data for farmers convenience.

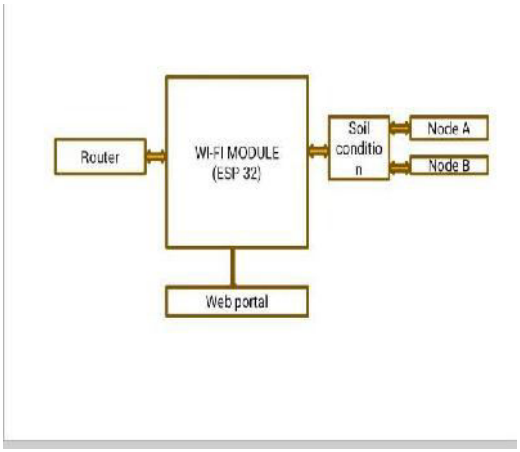


Fig , Block diagram of WI-FI section

ISBN 978-93-88808-61-3

Fig. ANIMAL ATTACK

MOTION DETECTOR : It will detect the thermal heat from animals body .GSM communication can be used to send immediate notification to farmer Alarm sounds can be activated.

DESCRIPTION: The AT-mega 328 is a low-power, high performance CMOS 28-bit microcontroller with 8K bytes of in-system programmable Flash memory. The device is manufactured using Atmel's high-density nonvolatile memory technology and is compatible with industry standard 80C51 instruction set and pin out. The Flash over on-chip allows the program memory to be reprogrammed in-system or by conventional non-volatile memory programmer. By combining a versatile 28-bit CPU with in-system programmable flash on a monolithic chip, the Atmel 328 is a powerful microcontroller which provides a highly flexible and cost effective solution to many embedded control applications. The analog to digital converter with 10 outputs has been connected to microcontroller which converts the analog data into digital format. The GSM sim 800 module along with WI-FI interconnect module that could store the data and send the information to the login channel to the farmers. In case of failure of the network connectivity immediately the GSM performs its operation by sending message to the farmers registered number. The combination of both i from the farm land internet and GSM could be performed based on its access and design over the data collection from the farm land.

Fig., GS M module to network connection

FUNTIONAL DESCRIPTION

WATER IRRIGATION

Water irrigation done through the basis of required to the plants and without wastage of the water. Such scientific method of water irrigation done by considering various parameters like soil type, crop type etc. The prevention of soil erosion practices which can drastically decrease negative effects associated with soil erosion such as reduced crop productivity, worsened water quality, lower effective reservoir water levels, flooding, and habitat destruction. Contour farming is considered an active form of sustainable agriculture.

SOFTWARE TOOLS

Software tools used

The software's which are used to developed this project are

- MicroKeil IDE compiler
- Languages used: Embedded C
- Things peak online web entry.

SQL Database and Power BI

At this point, the data found in the database needs to be transformed into a more user

represented as will not be understand SQL queries. Hence, to cater for this problem, Power BI is used to reconstruct the data into a visual representation such as a graph.

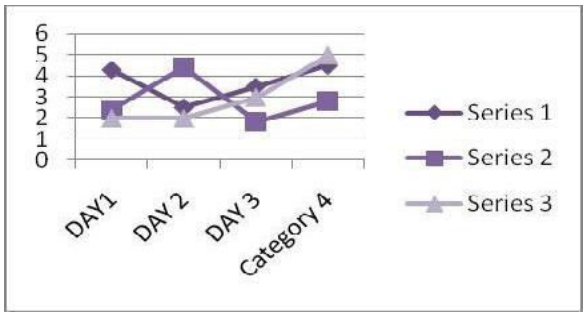


Fig. Graph Representation

Machine Learning

The Thing speak machine learning is the core logic of the proposed system. In general, a dataset is needed to train the machine to in the data in order to decide whether or not. For better precision, aOpenweathermap.com API is with the aim of knowing when the water pump needs to be opened. The pseudo code gives a simple illustration on how the machine learning system works producing code that is portable across wide platforms.

CONCLUSION

This multipurpose system gives an advance method to The system mainly monitors the behavior of soil moisture, air humidity, and air temperature and see how it contributes to evaluate the needs of water in a plant. The system uses machine learning and compares actual values obtained from sensors with a threshold value that has been fed to the machine learning for analysis. Next to this process, the machine learning cross checks the result obtained with weather forecast and then decides whether irrigation needs to be done or not. The farmer receives a notification on his smart phone and he can choose to turn on the water pump with a button click. Moreover, the system has a web app and is helpful if ever the farmer wants to see the statistical sensor data and assess the change in sensor readings throughout a time period. Furthermore, the system can calibrated for different type of plants, that is, the user is provided with a list of plants choices in his web app and mobile app. With this the user can choose the specific type of plant that is being cultivated and obtained threshold value and thus a more accurate irrigation prediction. Besides, an SMS system can be integrated if in case there is no internet connection. With this, the user would be notify about the prediction via an SMS and he can choose to switch on or off the water pump by replying to the SMS received. The entries that could also been saved safely for an farmers acknowledgement with date an time condition. The future works of transferring data is in the mode of social networks also through online data storage of cloud computation.

REFERENCES

1. *P. Narayut, P. Sasimanee, C.-I. Anupong, P. Phond and A. Khajonpong, 2016. A Control System in an Intelligent Farming by using Arduino Technology. Student Project Conference (ICT-ISPC), 2016 Fifth ICT International, pp. 53-56, 2016.*
2. *K. Benahmed, A. Douli, A. Bouzekri, M. Chabane and T. Benahmed, 2015. Smart Irrigation Using Internet of Things. Fourth International Conference on Future Generation Communication Technology (FGCT), 2015.*
3. *A.N. and K. D, 2016. Expe rimental investigation of remote control via Android smart phone of arduino based automated irrigation system using moisture sensor. 3rd International Conference on Electrical Energy Systems (ICEES), 2016.*

4. *T. Baranwal, N. and P. K. Pateriya, 2016. Development of IoT based Smart Security an d Monitoring Devices for Agriculture. 6th International Conference -Cloud System and Big Data Engineering (Confluence), 2016.*

5. *G. M.K., J. J. and A. M. G.S, 2015. Providing Smart Agriculture Solutions to Farmers for better yielding using IoT. IEEE Technological Innovation in ICT for Agriculture and Rural Development (TIAR), 2015.*

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

A SURVEY -EVENT DETECTION USING MACHINE LEARNING APPROACH IN CYBER-PHYSICAL SYSTEMS

[1]Dr.A.Kannagi
Professor, Department of CSE,
Malla Reddy College of Engineering,
kannagianbu_cse@mrce.in

[2]Mr.K.Praveen Reddy,
Assistant Professor, Department of CSE,
Malla Reddy College of Engineering,
praveenreddy_cse@mrce.in

Abstract

Asystematicapproach of extractingknowledgefromsensor data at various platforms plays a major role in the Data Mining Community to determine the intrusion detection in Cyber-Physical Systems (CPS), e.g., Assuming that there may be at whatever harm for building or aviation vehicles, those harm will be distinguished starting with the non stop arriving data. In the Existing methodology exhaustive Data Mining Framework which uses Differential Sensor Pattern (DSP) for Intrusion detection, DP miner has been used which greatly reduce the energy for calculation and correspondence in the CPS, the different pattern of sensors is been extracted that may have event information with a low communication cost but it can validate actual data only on lower data analysis whereas, for big data it cannot be sensed accurately. In order to achieve accuracy in big data environments, differential sensor mining technique with a machine-learning approach is been proposed for handling continuous quality improvement in event detection and it will useful for many CPS applications. .

KEYWORDS : Cyber-Physical Systems (CPS), Data Mining, Event Detection..

Introduction

Cyber-physical frameworks (CPSs) mix those learning What's more innovations of the third wave from claiming data processing, correspondence What's more registering with the learning Furthermore innovations of physical artifacts also engineered frameworks [1]. There appears with make a concurrence in the written works on the reality that those calling what's more learning of cyber physical frameworks would not monodisciplinary. However, it is at present debated if this discipline may be interdisciplinary, multi-disciplinary, and alternately trans-disciplinary to nature. Backers of the inter-disciplinary perspective contend that those mission fromclaiming CPSsscience Furthermore engineering organization is with make An span the middle of the two constituent learning domains, in particular the internet and the physical space. [2].

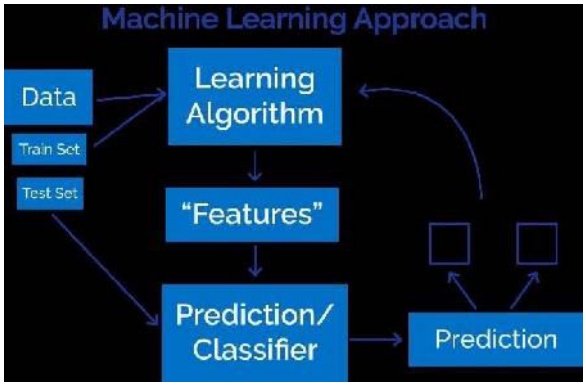
This argumentation appears on make right since majority of the data What's more communication science and technologies, on the person side, and physical framework science and technologies, on the other side, would epistemologically and methodologically different. The delegates of the multidisciplinary stance claim that the science technology about CPSs ought further synthesize the information and routines about the foundational physical, biological, building What's more data sciences, What's more if create a thorough science for CPSs. Those supporters of the trans-disciplinary elucidation case that once those science from claiming CPSs gives far extensive learning for implementation, the order if concentrate on giving requisition area free architectures What's more advances to fabricating useful cyber-physical artifacts and providing domain-orientated benefits [3].

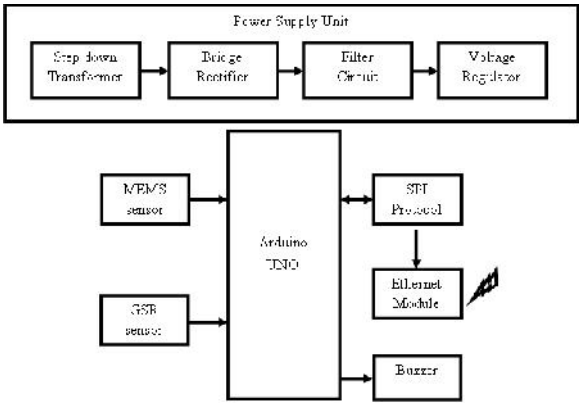
In our view, achieving all of these objectives can be considered as the mission of the science of CPSs. The objectives of the discipline of CPSs are:

- Mixing the information for different domains under a steady figure from claiming information with the goal as with underpin it perusing the fundamental standards about natural, formal, technical, social also human sciences.
- Creating a system-level understanding Furthermore theoretical frameworks from family of systems.

The principle Look into topics would for example, such that framework structure identification, advantageous interaction of physical and digital framework parts, combination from claiming empowering technologies, framework conduct analysis, self-sufficient system operation, ongoing framework control Also self-control, keen framework behavior, non-deterministic scenarios and protocols,

ISBN 978-93-88808-61-3





Also standards for next-generation usage. Likewise a whole, the order appears on a chance to be rather adolescent Also will be at present anguish starting with a to some degree unconsolidated, whether not confusing, wording. Test exploration On CPSs, and in addition prototyping-based testing would confronting experimental limits due to those vast scales, spatial distribution, inalienable complexity, prevailing heterogeneity and inserted nature. The idea and the term 'cyber-physical systems' popped dependent upon a percentage ten quite some time prior in the USA.

To Europe those same sort What's more manifestations for frameworks are named Possibly Likewise 'The Internet-of- Things'. [5], 'Web of Things', or as 'cooperative adaptive systems'.

Those expositive expression reflects An huge number for terms (such as, 'smart universal systems', 'deeply embedded systems', 'software-intensive systems', 'hybrid automata', sensor actuator networks, M2M (OECD), which attempt on indicate the same concept, setting accentuation once specific parts (Eg. Functionality, implementation, and applications) about complex frameworks that determinedly incorporate digital What's more physical parts. [6].

The utilization for different terms Eventually Tom's perusing Different scientists raises those inclination that they need aid working once totally distinctive field, However truth be told they deliver the same alternately fundamentally the same issues What's more aspects about CPSs. Hypothetical examination in this area about premium will beeven now really scattered What's more not streamlined. Actually, those expositive expression reveals to a huge number about models, also those mixture of reference frameworks. There would vast contrasts in the approaches, innovative work efforts, Also subsidizing projects in Europe, USA and Japan. The inspiration for our foundation investigate went starting with two perceptions. Our far reaching expositive expression study investigated that an expansive number about papers examines Also contributes to exactly particular parts about utilitarian frameworks, technologies, data flows, usage and requisitions of CPSs.

II. Methodology

A. Description

Machine learning algorithms differentiate into supervised or unsupervised. Supervised algorithms give both information also wanted output, furthermore with furnishing reaction something like those correctness about predictions throughout preparation. Over this, this procedure will apply which is nourished on new information. Unsupervised algorithms may be not necessity to prepare with fancied Conclusion information. Instead, they use an iterative methodology known as profound taking in will survey information also land at conclusions. Unsupervised learning algorithms only used for complex tasks than supervised learning systems.

The methods included over machine learning comparative to that of data mining and predictive modelling. Both obliges seeking through information will search for patterns and changing programme actions appropriately. Numerous individuals need aid great known In light of the use for machine learning in starting with shopping on the web Furthermore being served ads identified with their buy. This

ISBN 978-93-88808-61-3

REFERENCE PAPER	DESCRIPTION	ALGORITHMS USED	EVALUATION
Adaptive Layered Approach using Machine Learning Techniques with Gain Ratio for Intrusion Detection Systems	• In this paper, An multi-layer interution detection model will be planned and created to attain effectiveness which improves the detection and classification rate accuracy. • Machine learning techniques (C5 decision tree, Multilayer Perceptron neural system Also Naïve Bayes) need been connected utilizing gain ratio to selecting the best Characteristics for each layer to utilize smaller storage space and get higher intrusion detection.	Machine Learning Approach like C5 decision tree, Multilayer Perceptron neural network and Naïve Bayes are used which produces High intrusion performance.	Only C5 eliminates False alarm rate for MLP. other algorithms face difficulty to eliminate False alarm rate for MLP
Machine learning-based CPS for clustering high throughputmachining cycle conditions	• In this paper, unsupervised machine learning algorithms in cyber-physical systems are the key features to work towards highly precise diagnosis tools. • In case of clustering techniques, the Gaussian mixture model is used and also provides optimal solution in terms of interpretation by machine tool experts. • The agglomerative hierarchical algorithm is used to determine cycle phases. K means as same as agglomerative hierarchical algorithm to inherit variables.	Gaussian mixture model. The agglomerative algorithm and K-means clustering are used.	No importance shown to upgrade CPS embedded electronics which enables the algorithm to implement on its FPGA.
Event Detection through Differential PatternMining in Cyber-Physical Systems	• In this paper, DP miner, an exhaustive data mining framework for using in wireless sensors which performs in a distributed and parallel manner and it is able to extract a pattern of sensors that have event information. • DPminer can greatly reduce the energy for computation and communication in the Cyber Physical Systems.	DP miner and Differential Sensor Mining technique is used.	Differential Sensor Mining Technique faces difficulty in analysing big data in MLP.

happens in light suggestion engines utilize machine learning in to identity test web promotion conveyance previously, very nearly constant. Likewise separated starting with the customize marketing, other as a relatable point machine learning in employments instances incorporate duplicity detection, spam filtering, system security risk detection, predictive upkeep Also fabricating news encourages.

Fig1. Block Diagram of Machine Learning Approach. For (supervised) classification and regression (the most common tasks):

- Algorithm selection: Choose an algorithm.
- Feature selection: Choose features that capture the important characteristics of the system.
- Training/model building: Use part of the labeled set to build the model
- Parameter optimization (cross validation): Optimize the parameters using a second part of the labeled set to minimize the error rate.
- Validation: Use the remainder of the dataset to validate and assess the performance of the tuned model
- Apply the Algorithm

EXAMPLE: "MEDICAL MONITORING- POST OPERATIVE WOUND ANALYTICS"

Patients after an operation usually go through the recovery/rehabilitation process where they follow a strict routine. That will do by using sensors.

After the major surgery as per instruction from surgerion patients should maintain a fixed position or else if the patients supposed to falls down. That the position level will be monitored (MEMS Sensor) .MEMS generally consists three position like x, y, z. If the changes will be in the position means that will be updated through web server. Because of this updating nurses or Ward in charge can get alert without direct monitoring.

GSR, standing for galvanic skin response, is a method of measuring the electrical conductance of the skin. Strong emotion can cause stimulus to your sympathetic nervous system. Due to this condition can able to know the Pain or stress level (GSR Sensor) which rose after involved in surgery will be viewed through the web page. Not only web page updation can give alert through buzzer also. If the sensor data is not received to the cloud means

the doctor or representative person cannot able to monitoring the patient health frequently. So that patient can be affected by unwanted pain or any other factors. So that our machine-learning approach will guide to rectify/ notify the problem like sensor failure, controller board failure, internet connection lost.

Conclusion

Thus in this survey we analyze several algorithms based on Machine Learning Approach in order to extract knowledge from sensor data at various platforms which performs critical piece in the data mining to figure out those occasion identification to Cyber-Physical frameworks (CPS) contrasting with differential sensor pattern (DSP) .we use machine learning algorithm for event detection where we implement C5, Decision Tree mining techniques etc where accurate predictive results are achieved and Anomaly detection, Gaussian Mixture model, agglomerative hierarchical algorithm and K-means clustering is surveyed for supporting big data analysis.

References

1. "Adaptive Layered Approach using Machine Learning Techniques with Gain Ratio for Intrusion Detection Systems", Heba Ezzat Ibrahim etal International Journal of Computer Applications (0975 – 8887) Volume 56– No.7, October 2012
2. "Machine learning-based CPS for clustering high throughput machining cycle conditions", Javier Diaz-Rozoa,b*, Concha Bielzab, Pedro Larrañagab open access article 2017
3. "Event Detection through Differential PatternMining in Cyber-Physical Systems", Md Zakirul Alam Bhuiyan, Member, IEEE, Jie Wu, IEEE transactions on big data,2017
4. B. Sharma, I. Franjo, N.-M. Alexandru, C. Haifeng, and. Guofei, "Modeling and analytics for cyber-physical systems in the age of big data," SIGMETRICS Perform.
5. Evaluation Review,vol. 41, no. 4, pp. 74–77, 2014.

6. G. Spezzano and A. Vinci, "Pattern detection in cyber physical systems," Procedia Computer Science, vol. 52, no. 2015, pp. 1016–1021, 2015.O. Niggemann1, G. Biswas, J. S. Kinnebrew, H. Khorasgani,S. Volgmann, and A. Bunte,
7. "Data-driven monitoring of cyber physical systems leveraging on big data and the internet-of things for diagnosis and control," in Proceedings of the 26th 1International Workshop on Principles of Diagnosis, 2015, pp. 185–192.
8. J. Lee, H. D. Ardakani, S. Yang, and B. Bagheri, "Industrial big data analytics and cyber-physical systems for future maintenance and service innovation," Procedia CIRP, vol. 38, pp. 3–7, 2015.
9. J. Jara, D. Genoud, and Y. Bocchi, "Big data for cyber physical systems: An analysis of challenges, solutions and opportunities," in The 8th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2014, pp. 376–380.
10. Jensen, T. D., Jensen, F.V., and Nielsen. (2001). "Bayesian networks and decision graphs," Springer, Berlin.
11. Amor, N. B., Benferhat, S., and Elouedi, Z. (2003). Naive Bayesian networks in intrusion detection systems. In Proc. Workshop on Probabilistic Graphical Models for Classification, 14th European Conference on Machine Learning (ECML) and the 7th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD),Croatia.
12. Hall, M. A. (1999). "Correlation based Feature Selection for Machine Learning,". Available at: https://www.lri.fr/pierres/ donn% E9es/save/these/articles/lprqueue/ hall99correlationbased.pdf
13. Draper-Gil, G., Lashkari,A. H., Mamun, M.S. I., and Ghorbani, A. A.(2017). "Characterization of Encrypted and VPN Traffic using Tim erelated Features," in Proceedings of the 2nd International Conference on Information Systems Security and Privacy (ICISSP 2016), 407–414.

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

A Review on Data Preprocessing For Efficient Prediction in Customer Relationship Management

[1]Ganga Patur, Assistant Professor.
[2]D. Nilima Priyadarshini, Assistant Professor.
Malla Reddy College of Engineering,

ABSTRCT

CRM (ConsumerRelationship Management) is a customer-focused business strategy designed to optimize revenue, profitability, and customer loyalty. CRM can use information from outside or within a company allowing much better comprehension of its customers on the set basis or to your own foundation, by producing client personalized documents. An improved knowledge of the buyer's customs, pursuits and demands might grow the transaction. So, steady information regarding your clients' choices and preferences forms the cornerstone of productive CRM. Since organizations become internet (in other words, grow in to e business), the find it difficult to maintain faithfulness in their older customers and also to entice clients remains more crucial, as a competitor's enterprise internet site might be only 1 click away. In this paper we studied data prepossessing methods for client log data.

Keywords: Data prepossessing, log, competitor prediction and Big data.

INTRODUCTION

Voluminous of information active in those on-line World Wide Web have managed to get rather vital that you utilize automatic data mining and knowledge discovery procedures to learn person navigation tastes. The various manners of internet website usage using way of a specific user could possibly be detected with World Wide Web usage mining methods that can mechanically recover ordinary accessibility patterns employing the utilization of sooner user simply click flows utilized in weblog data files. These Programs might be properly used towards designing the internet page for your own user and also to encourage digital advertising. Net usage mining technologies incorporates methods from two hot search areas, specifically, data mining and also the World Wide Web. By assessing the competition understanding concealed in blogs,

internet usage mining may assist searchers to supply much better layout and enterprise worries to present much better navigation behavior. Many businesses are emphasizing buyer orientation to both maintain regular users to its growth of consumer relationship administration. Investigation of curious browsers, gives invaluable advice for internet site designer to swiftly react for their own unique wants. This chapter introduces the search methodology utilized to look exactly the upcoming page forecast approach.

CUSTOMER LOGDATA

Purchaser log info can be really a document that has tremendous sum of facts and by that data origin; lots of info abstractions might be generated. For example, page opinions, host periods, along with click-streams. In these abstracts, shared provisions and key words can be utilized as specified in Table 1. This portion in addition supplies an in-depth outline of this web-log document structure used from today's research work.

A log file will be understood to be a document which enrolls the surgeries of the internet server. Log data files returns advice such as for instance the data files which can be asked, sometime of this document ask the individual and also the speaking webpage. Every point of this log document defines one “strike" over the log file from your host plus comprises numerous subjects and also the arrangement of this log utilized for assesses change from host to host. Investigation of log document is Deemed valuable for the next reasons:

• The Internet server produces log documents, therefore getting an raw info is Not Too hard and Doesn't require any alterations or added programming attempt,

• Business's servers may maintain info inside their standard. It makes it possible to get a Institution to Alter applications after, utilize a lot Diverse applications and analyze chronological arrangement having a new program,

• Production and incorporating details for the log record doesn't need any extra Domain Name Server Look-ups. Ergo, There Aren't Any external server requirements which may slow down page loading rates, also Contributes to uncounted webpage viewpoints, and also

• The Internet Website's host documents all Trade it gets and this is considered reliable.

The arrangement of this log record is displayed at Table 1 & 2An hyphen ('-')

at one or more of these disciplines suggest missing info.

TABLE 1 IMPORTANT TERMS IN CUSTOMER LOG DATA

S.No.	Terms	Description
1	User	**Users accessing file from the web servers through a browser.**
2	Page View	**A page view is an abstract that consist of every file that is displayed on user's browser screen at one point of time. A page view may be associated with a single user action or can be related with several files such as scripts ,frames,and graphics, etc.,**
3	Hit	**Every successful file that is sent to the web browser is a hit**
4	Click Stream	**It is a sequential series of page view requests.**
5	Server Session or visit	**A Server Session or visit happens when a user or robot visits a website.**
6	User Session	**A user session is defined as a set of page requests made by a single user.**
7	**Customer Log**	**These are files that stores into them details regarding all the visits made to a web site or a portal automatically and are maintained in the web server.**

TABLE 2 CUSTOMER LOG FIL

S.No.	Name of Field	Description	Example value
1	IP Address	IP address of the Client who request for a page on the web server	**127.0.0.1**
2	UserID and Password	Provides the username and their corresponding password used during the access of a content-secured transaction	**Voder23 12ert35**
3	Timestamp	The date, time and time zone when the server finished processing the request.	**[10/Oct/2000:13:55:36 -0700]**
4	Access Request	Request line from the client. It has three parts, the METHOD, URL STEM and PROTOCOL used during transmission.	**GET http/www.yahoo.com/asctab31.zipHTTP/1.0**
	Method	Can be GET (request made to get a program or document) or POST (during transmission indicates the server that data is following) or HEAD (used by link checking programs, not browsers and downloads just the information in the HEAD tag information)	**GET POST HEAD**
	URL	The address of	**/download/windows/asctab31.zip**
	Protocol	**protocol**	**HTTP/1.0**

I. RESEARCH METHODOLOGY

Internet can be actually a client/server style and design by which a consumer sends an internet requests for within the web (WWW) into some internet server. The internet server reacts by reacting to this petition. The trade session includes the market of protocols and methods. But as a result of exponential increase of WWW, there really are a high quantity of customers that disagrees with all the servers with a high number of programs correlated with just one another, causing a significant raise the WWW latency and burden about the internet. If a proxy host set in between a web browser and a host, it's a effective tool which could possibly be utilized to decrease your WWW's latency. It follows that it may intercept any orders into the server to guarantee whether the request can be fulfilled by the client itself. If not, then it may be offered to the internet server. The clear presence of proxy servers also provides 2 major positive aspects as supplied just below.

- Reduce latency: Gradually, most of the asked consequences from several customers are saved inside a proxy-server.For example, contemplate if just two users and B get the web by means of a proxy host. Assume consumer A asks to get a specific webpage (P 1). Shortly after, consumer also requests for equal webpage. Without forwarding the petition for the internet server, then these pages is returned from your proxy host its own cache at which in fact the newly downloaded website pages have been kept. Considering proxy host and the consumer share exactly the Exact Same system, the surgeries are substantially quicker, thereby decreasing the perceived latency, and also
- Filter un-wanted Requests: Negative asks are all taken off from the Proxy servers. By way of instance, a faculty can confine the college students from obtaining a particular pair of the web sites using a proxy-server.

To reduce the WWW latency, the behavior of the consumer can be called and therefore the pages that are predicted are all pre-fetched and kept temporarily at the cache from their proxy host. The petition of this user could be fulfilled immediately when the webpage can be found from your cache. An overall site forecast version is displayed in Figure.1.

Web Requests

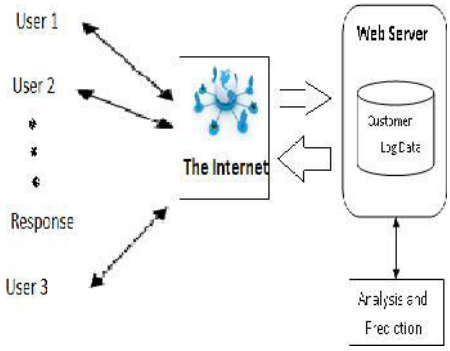


Figure1: General Architecture of Web System with Web Access Prediction

Due to person's successive actions within their communicating with the entire Internet presents a huge obstacle for investigators from the internet engineering field and also can be the primary focus with the exploration, forecast of person's foreseeable future asks is composed of varied endeavors and determine fig 1 offer the stream of those activities at the research job. The suggested strategy is known as adjoining page forecast approach. This job includes three major actions, particularly, pre-processing, competition consumer identification along with forecast of all future asks.

Inside this exploration function, every one of the aforementioned ways is taken care of like an individual period, which must be implemented at a sequential way throughout the plan and execution of internet site forecast procedure. The investigation methodology has been intended in a fashion that all measure tries to increase its individual endeavor and operates with all the intention of bettering its performance prediction. Throughout the stream of forecast, the outcome of one particular phase can be utilized as input signal the subsequent period. The suggested research frame is offered in Figure 3.3 along with the many processes enhanced throughout the plan of the next page forecast approach are all introduced at these sub sections.

Phase I: Prepossessing Algorithms

Prepossessing of a web log file is nothing but simply reformatting the entries of a log file into a form that can be used directly by the subsequent steps of the log analyzer.

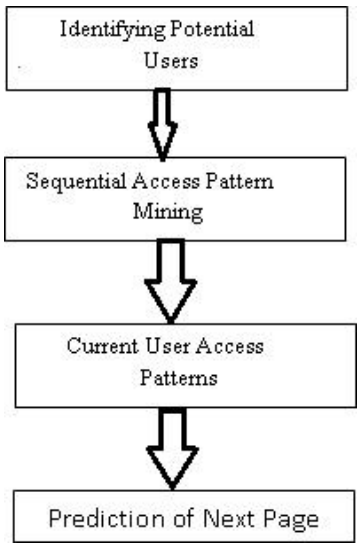


Figure 3.2: Tasks in next page prediction system

II. PREPROCESSING ALGORITHMS

The very first thing of this suggested second web-page forecast process will be per-processing, at which in fact the most important focus would be always to maintain simply applicable data out of the uncooked link. As a result of great number of insignificant data while in the internet log, the log may not be specifically utilized from the internet log mining treatment, thus at the prepossessing period, uncooked Internet logs will need to get cleaned, examined and changed to additional usage.

Period I of this analysis plays per-processing in 5 actions. They're recorded below and also the processes utilized in every measure are explained inside this chapter.

- Cleaning,
- User identification,
- Session identification,
- Formatting, and
- Clustering

III. CLEANING CUSTOMER LOG DATA

In the first step, that is, the task of cleaning raw web log data is considered. The data removed during cleaning are not required for user navigation and hence can be deleted safely from the log file. This step carries out the following tasks:-

- Removal of unwanted and redundant data,
- Removal of non-human accesses, and
- Removal of erroneous references.

Cases of undesirable data comprise asks including graphics, java script sand flash cartoons and video clip, etc. In case the file name contains gif, jpg, JPEG, CSS and so forth they are pruned out from the internet log document. Redundant statistics are recordings using similar

values in every single characteristic of this report. Instance of these statistics comprises admissions created by webmasters along with Spider accesses (instruments that scanning internet site to automatically extract its content). Search Engine normally utilizes system bots to creep throughout the web pages to get advice. The amount of information generated with these robots at a log record is high and has got a very poor impact whilst detecting navigation layout. This issue is solved inside thispaper by pinpointing the exact robot entrances first prior to devoting an individual collection in to rival and not-competitor end users. As stated by entrances from web-log produced with system robots can be identified by their IP address and agents. But this might require comprehension on most of form of representatives and see's, and this isn't easy to have. Another method will be to review the robots.txt document (positioned in the site's root directory), since a system convention has to read this document before obtaining the site. This really is due to the fact that the robots.txt gets got the access information of the site and every single robot will petition to learn its accessibility before scrawling. But that can't be relied on since obedience with robot exclusion standard is voluntary & the majority of the bots usually do not comply with exactly the suggested benchmark. So, to manually delete custom entrances, the next treatment issued.

Detect and remove all entries which has accessed robots.txtfile

Detect and remove all entries with visiting time of access as midnight (commonly used as the network activity at that time is light)

Remove entry when access mode is HEAD instead of GET orPOST

Compute browsing speed and remove all entries whose speed less than two seconds. The browsing speed is calculated as the number of viewed pages / session time.

CONCLUSION

In this paper we studied competitor prediction, in order to this first data per-processing is required, Real world data are generally Incomplete, Noisy and Inconsistent. Data cleaning, also called data cleansing or scrubbing. Fill in missing values, smooth noisy data, identify or remove the outliers, and resolve inconsistencies.Data cleaning is required because source systems contain "dirty data" that must be cleaned.In a customer relationship management (CRM) context, data prepossessing is a component of Web mining. Web usage logs may be pre-processed to extract

meaningful sets of data called user transactions, which consist of groups of URL references. User sessions may be tracked to identify the user, the Web sites requested and their order, and the length of time spent on each one. Once these have been pulled out of the raw data, they yield more useful information that can be put to the user's purposes, such as consumer research, marketing, or prediction.

REFERENCES

1. Ashish Bindra ;SrinivasuluPokuri ; Krishna Uppala ; Ankur Teredesai, 2012, "Distributed Big Advertiser Data Mining", ISSN: 2375-9232, 2012 IEEE 12th International Conference on Data Mining Workshops, PP: 914-914.
2. Abdel-Karim Al-Tamimi ; Raj Jain ; Chakchai So-In, 2010, "Dynamic resource allocation based on online traffic prediction for video streams", 2010 IEEE 4th International Conference on Internet Multimedia Services Architecture and Application, PP: 1---6.
3. BenleSu ;Yumei Wang ; Yu Liu, 2016, "Analysis and prediction of content popularity for online video service: a Youku case study", ISSN: 1673-5447, Volume: 13 , Issue: 12 , PP: 216-233.
4. ChengangZhu ;Guang Cheng ; Kun Wang, 2017, "Big Data Analytics for Program Popularity Prediction in Broadcast TV Industries", ISSN: 2169-3536, Volume: 5, PP: 24593-24601.
5. David K. Becker, 2017, "Predicting outcomes for big data projects: Big Data Project Dynamics (BDPD): Research in progress", 2017 IEEE International Conference on Big Data (Big Data), PP: 2320-2330.
6. Jian Ming ;Lingling Zhang ; Jinhai Sun ; Yi Zhang, 2018, "Analysis models of technical and economic data of mining enterprises based on big data analysis", 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), PP: 224-227.
7. Katarina Grolinger ; Miriam A.M. Capretz ; Luke Seewald, 2016, "Energy Consumption Prediction with Big Data: Balancing Prediction Accuracy and Computational Resources", 2016 IEEE International Congress on Big Data (BigData Congress), PP: 157-164.

8. KunZhang ;Minrui Fei ; Jianguo Wu ; Peijian Zhang, 2013, “Fast prediction model based big data system identification”, 2013 Chinese Automation Congress, PP: 465-469.
9. Pedro Bastos ; Rui Lopes ; Luís Pires ; Tiago Pedrosa, 2009, “Maintenance behaviour-based prediction system using data mining”, ISSN: 2157-3611, 2009 IEEE International Conference on Industrial Engineering and Engineering Management, PP: 2487-2491.
10. XiaojingMa ;Zhitang Li ; Hao Tu ; Bochao Zhang, 2010, “A Data Hiding Algorithm for H.264/AVC Video Streams Without Intra- Frame Distortion Drift”, ISSN: 1051-8215, Volume: 20 , Issue: 10 , PP: 1320-1330.

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

A Routing Protocol for Enhanced Efficiency in TEEN

Kota Divya Bharathi[1], Thirumala Vasala[2]
Assistant Professors, CSE Department,
Malla Reddy College of Engineering, Dhulapally, Kompally, Secunderabad-500100

Abstract

Wireless sensor networks are expected to find wide applicability and increasing deployment in the near future. In this paper, we propose a formal classification of sensor net- works, based on their mode of functioning, as proactive and reactive networks. Reactive networks, as opposed to passive data collecting proactive networks, respond immediately to changes in the relevant parameters of interest. We also in troduce a new energy efficient protocol, TEEN (Threshold sensitive Energy Efficient sensor Network protocol) for re- active networks. We evaluate the performance of our proto- col for a simple temperature sensing application. In terms of energy efficiency, our protocol has been observed to out- perform existing conventional sensornetwork protocols.

1. Introduction

In recent years, the use of wired sensor networks is being advocated for a number of applications. Some examples include distribution of thousands of sensors and wires over strategic locations in a structure such as an airplane, so that conditions can be constantly monitored both from the inside and the outside and a real-time warning can be issued when the monitored structure is about to fail.

Sensor networks are usually unattended and need to be fault-tolerant so that the need for maintenance is minimized. This is especially desirable in those applications where the sensors may be embedded in the structure or are in inhospitable terrain and are inaccessible for any service. The advancement in technology has made it possible to have extremely small, low powered devices equipped with programmable computing, multiple parameter sensing and wireless communication capability. Also, the low cost of sensors makes it possible to have a network of hundreds or thousands of these wireless sensors, thereby enhancing the reliability and accuracy of data and the area coverage as well. Also, it is necessary that the sensors be easy to deploy (i.e., require no installation cost etc). Protocols for these networks must be designed in such a way that the limited power in the sensor nodes is efficiently used. In addition, environments in which these nodes operate and respond are very dynamic, with fast changing physical parameters. The following are some of the parameters which might change dynamically depending on the application:

Power availability.

Position (if the nodes are mobile).

Reachability.

Type of task (i.e. attributes the nodes need to operate on)

So, the routing protocol should be fault-tolerant in such a dynamic environment. The traditional routing protocols defined for wireless ad hoc networks [1] [9] are not well suited due to the following reasons:

1. Sensor networks are "data centric" i.e., unlike traditionalnetworkswheredataisrequestedfromaspecific node, datais requestedbasedoncertainattributessuch as, which area has temperature $> 50\ F$?
2. The requirements of the network change with the application and so, it is application-specific [3]. For example, in some applications the sensor nodes are fixed and not mobile, while others need data based only on one attribute (i.e., attribute is fixedin this network).
3. Adjacent nodes may have similar data. So, rather than sending data separately from each node to therequesting node, it is desirable to aggregate similar data and send it.
4. In traditional wired and wireless networks, each node is given a unique id, used for routing. This cannot be effectively used in sensor networks. This is because, these networks being data centric, routing to and from specific nodes is not required. Also, the large number of nodes in the network implies large ids [2], which might be substantially larger than the actual data being transmitted.

This work is supported by the Ohio Board of Regents' Doctoral En- hancement Funds

Thus, sensor networks need protocols which are application specific, data centric, capable of aggregating data and optimizing energy consumption. An ideal sensor network should have the following additional features:

Attribute based addressing is typically employed in sensor networks. The attribute based addresses are composed of a series of attribute-value pairs which specify certain physical parameters to be sensed. For example, an attribute address may be (temperature *> 100 F* , location = ??). So, all nodes which sense a temperature greater than 100 *F* should respond with their location.

Location awareness is another important issue. Since most data collection is based on location, it is desirable that the nodes know their position whenever needed.

2. Related Work

In this section, we provide a brief overview of some related research work.

Intanagonwiwat et. al [7] have introduced a data dissemination paradigm called *directed diffusion* for sensor networks. It is a data-centric paradigm and its application to query dissemination and processing has been demonstrated in this work.

Estrin et. al [3] discuss a hierarchical clustering method with emphasis on localized behavior and the need for asymmetric communication and energy conservation in sensor networks.

A cluster based routing protocol (CBRP) has been proposed by Jiang et. al in [8] for mobile ad-hoc networks. It divides the network nodes into a number of overlapping or disjoint two-hop-diameter clusters in a distributed manner. However, this protocol is not suitable for energy constrained sensor networks in this form.

Heinzelman et. al [5] introduce a hierarchical clustering algorithm for sensor networks, called *LEACH*. We discuss this in greater detail in section 6.1.

3. Motivation

In the current body of research done in the area of wireless sensor networks, we see that particular attention has not been given to the time criticality of the target applications. Most current protocols assume a sensor network collecting data periodically from its environment or responding to a particular query. We feel that there exists a need for networks geared towards responding immediately to changes in the sensed attributes. We also believe that sensor networks should provide the end user with the ability to control the trade-off between energy efficiency, accuracy and response times dynamically. So, in our research, we have focussed on developing a communication protocol which can fulfill these requirements.

4. Classification of Sensor Networks

Here, we present a simple classification of sensor networks on the basis of their mode of functioning and the type of target application.

Proactive Networks

The nodes in this network periodically switch on their sensors and transmitters, sense the environment and transmit the data of interest. Thus, they provide a snapshot of the relevant parameters at regular intervals. They are well suited for applications requiring periodic data monitoring.

Reactive Networks

In this scheme the nodes react immediately to sudden and drastic changes in the value of a sensed attribute. As such, they are well suited for time critical applications.

5. Sensor Network Model

We now consider a model which is well suited for these sensor networks. It is based on the model developed by Heinzelman et. al. in [5]. It consists of a base station*(BS)*, away from the nodes, through which the end user can access data from the sensor network. All the nodes in the network are homogeneous and begin with the same initial energy. The *BS* however has a constant power supply and so, has no energy constraints. It can transmit with high power to all the nodes. Thus, there is no need for routing from the *BS* to any specific node. However, the nodes cannot always reply to the *BS* directly due to their power constraints, resulting in asymmetric communication.

This model uses a hierarchical clustering scheme. Consider the partial network structure shown in Fig. 1. Each cluster has a cluster head which collects data from its cluster members, aggregates it and sends it to the *BS* or an upper level cluster head. For example, nodes 1.1.1, 1.1.2, 1.1.3, 1.1.4, 1.1.5 and 1.1 form a cluster with node 1.1 as the cluster head. Similarly there exist other cluster heads such as 1.2, 1 etc. These cluster-heads, in turn, form a cluster with node 1 as their cluster-head. So, node 1 becomes a second level cluster head too. This pattern is repeated to form a hierarchy of clusters with the uppermost level cluster nodes reporting directly to the *BS*. The *BS* forms the root of this hierarchy and supervises the entire network. The main features of such an architecture are:

All the nodes need to transmit only to their immediate cluster-head, thus saving energy.

Only the cluster head needs to perform additional computations on the data. So, energy is again conserved.

ISBN 978-93-88808-61-3

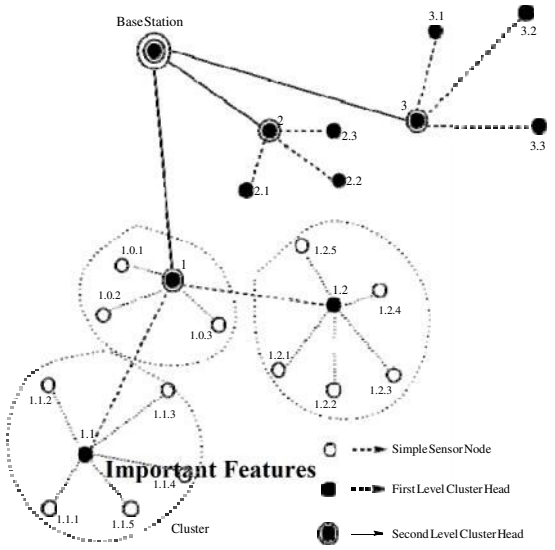


Figure 1. *HierarchicalClustering*

Cluster-heads at increasing levels in the hierarchy need to transmit data over correspondingly larger distances. Combined with the extra computations they perform, they end up consuming energy faster than the other nodes. In order to evenly distribute this consumption, all the nodes take turns becoming the cluster head for a time interval *T*, called the cluster period.

6. Sensor Network Protocols

The sensor network modeldescribed in section 5 is used extensively in the following discussion of sensor network protocols.

Proactive Network Protocol

In this section, we discuss the functionality and the characteristics expected in a protocol for proactivenetworks.

Functioning

At each cluster change time, once the cluster-heads are decided, the cluster-head broadcasts the following parameters :

***Report Time(T_R)*:** This is the time period between succes- sive reports sent by a node.

***Attributes(A)*:** This is a set of physical parameters which the user is interested in obtaining data about.

At every report time, the cluster members sense the parameters specified in the attributes and send the data to the cluster-head. The cluster-head aggregates this data and sends it to the base station or the higher level cluster-head, as the case may be. This ensures that the user has a com- plete picture of the entire area covered by the network.

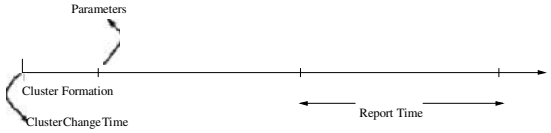


Figure 2. *Time line for proactive protocol*

The important features of this scheme are mentioned below:

1. Since the nodes switch off their sensors and transmitters at all times except the report times, the energy of the network is conserved.

2. At every cluster change time, T_R and A are transmitted afresh and so, can be changed. Thus, the user can decide what parameters to sense and how often to sense them by changing A and T_R respectively.

This scheme, however, has an important drawback. Because of the periodicity with which the data is sensed, it is possible that time critical data may reach the user only after the report time. Thus, this scheme may not be very suitable for time-critical data sensing applications.

LEACH

LEACH (Low-Energy Adaptive Clustering Hierarchy) is a family of protocols developed in [5]. LEACH is a good approximation of a proactive network protocol, with some minor differences.

Once the clusters are formed, the cluster heads broadcast a TDMA schedule giving the order in which the cluster members can transmit their data. The total time required to complete this schedule is called the frame time T_F . Every node in the cluster has its own slot in the frame, during which it transmits data to the cluster head. When the last node in the schedule has transmitted its data, the schedule repeats.

The *report time* discussed earlier is equivalent to the *frame time* in LEACH. The *frame time* is not broadcast by the cluster head, though it is derived from the TDMA schedule. However, it is not under user control. Also, the attributes are predetermined and are not changed midway.

Example Applications

This network can be used to monitor machinery for fault detection and diagnosis. It can also be used to collect data about temperature change patterns over a particular area.

Reactive Network Protocol: TEEN

In this section, we present a new network protocol called TEEN *(**T**hreshold sensitive **E**nergy **E**fficient sensor **N**etwork protocol)*. It is targeted at reactive networks and is the first protocol developed for reactive networks, to our knowledge.

Functioning

In this scheme, at every cluster change time, in addition to the attributes, the cluster-head broadcasts to its members,

Hard Threshold ***(H_T)*****:** This is a threshold value for the sensed attribute. It is the absolute value of the attribute beyond which, the node sensing this value must switch on its transmitter and report to its clusterhead.

Soft Threshold ***(S_T)*****:** This is a small change in the value of the sensed attribute which triggers the node to switch on its transmitter and transmit.

The nodes sense their environment continuously. The first time a parameter from the attribute set reaches its hard threshold value, the node switches on its transmitter and sends the sensed data. The sensed value is stored in an internal variable in the node, called the *sensed value (SV)*. The nodes will next transmit data in the current cluster period, only when *both* the following conditions aretrue:

1. The current value of the sensed attribute is greater than the hard threshold.

2. The current value of the sensed attribute differs from *SV* by an amount equal to or greater than the soft threshold.

Whenever a node transmits data, *SV* is set equal to the current value of the sensed attribute.

Thus, the hard threshold tries to reduce the number of transmissions by allowing the nodes to transmit only when the sensed attribute is in the range of interest. The soft threshold further reduces the number of transmissions by eliminating all the transmissions which might have otherwise occurred when there is little or no change in the sensed attribute once the hard threshold.

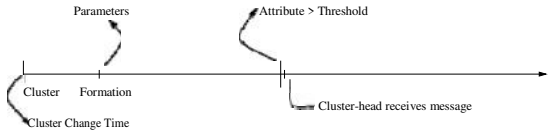


Figure 3. *Time Line for TEEN*

Important Features

The main features of this scheme are as follows:

1. Time critical data reaches the user almost instantaneously. So, this scheme is eminently suited for time-critical data sensing applications.

2. Message transmission consumes much more energy than data sensing. So, even though the nodes sense continuously, the energy consumption in this scheme can potentially be much less than in the proactive network, because data transmission is done less frequently.

3. The soft threshold can be varied, depending on the criticality of the sensed attribute and the target application.

4. A smaller value of the soft threshold gives a more accurate picture of the network, at the expense of increased energy consumption. Thus, the user can control the trade-off between energy efficiency and accuracy.

5. At every cluster change time, the attributes are broadcast afresh and so, the user can change them as required.

The main drawback of this scheme is that, if the thresholds are not reached, the nodes will never communicate, the user will not get any data from the network at all and will not come to know even if all the nodes die. Thus, this scheme is not well suited for applications where the user needs to get data on a regular basis. Another possible problem with this scheme is that a practical implementation would have to ensure that there are no collisions in the cluster. TDMA scheduling of the nodes can be used to avoid this problem. This will however introduce a delay in the reporting of the time-critical data. CDMA is anotherpossible solution to this problem.

Example Applications

This protocol is best suited for time critical applications such as intrusion detection, explosion detection etc.

7. Performance Evaluation

Simulation

To evaluate the performance of our protocol, we have implemented it on the ns-2 simulator [10] with the *LEACH* extension [4]. Our goals in conducting the simulation are as follows:

Compare the performance of the TEEN and LEACH protocols on the basis of energy dissipation and the longevity of the network.

Study the effect of the soft threshold s_T on TEEN.

The simulation has been performed on a network of 100 nodes and a fixed base station. The nodes are placed randomly in the network. All the nodes start with an initial energy of 2J. Cluster formation is done as in the *leach* protocol [5] [6]. However, their radio model is modified to include idle time power dissipation (set equal to the radio electronics energy) and sensing power dissipation (set equal to 10% of the radio electronics energy). The idle time power is the same for all the networks and hence, does not affect the performance comparison of the protocols.

Simulated Environment

For our experiments, we simulated an environment with varying temperature in different regions. The sensor network nodes are first placed randomly in a bounding area of 100x100 units. The actual area covered by the network is then divided into four quadrants. Each quadrant is later assigned a random temperature between $0\ F$ and $200\ F$ every 5 seconds during the simulations. It is observed that most of the clusters have been well distributed over the four quadrants.

Experiments

We use two metrics to analyze and compare the performance of the protocols. They are:

Average energy dissipated: This metric shows the average dissipation of energy per node over time in the network as it performs various functions such as transmitting, receiving, sensing, aggregation of data etc.

Total number of nodes alive: This metric indicates the overall lifetime of the network. More importantly, it gives an idea of the area coverage of the network over time.

We now look at the various parameters used in the implementation of these protocols. A common parameter for both the protocols is the attribute to be sensed, which is the temperature.

The performance of TEEN is studied in two modes, one with only the hard threshold *(hard mode)* and the other with both the hard threshold and the soft threshold *(soft mode)*. The hard threshold is set at the average value of the lowest and the highest possible temperatures, $100\ F$. The soft threshold is set at $2\ F$ for our experiments.

Results

We executed 5 runs of the simulator for each protocol and for each mode of TEEN. The readings from these 5 trials were then averaged and plotted. A lower value of the energy-dissipation metric and a higher number of nodes alive at any given time indicates a more efficient protocol.

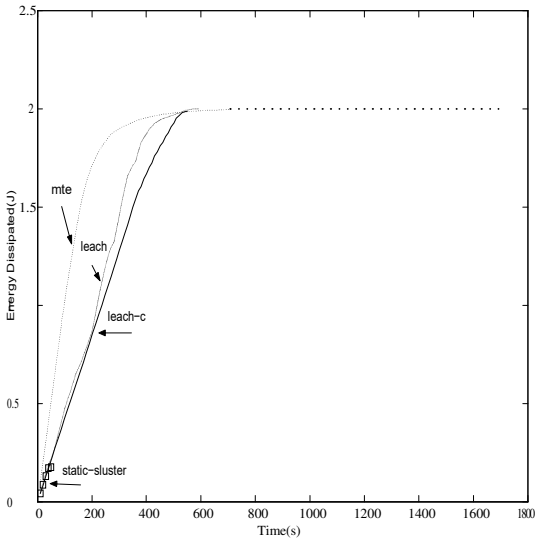


Figure 4. *Energy dissipation: LEACH*

Figures 4 and 5 show the behavior of the network in proactive mode. This comparison was originally done in LEACH [6]. It is repeated here taking into account the modified radio energy model. Of the four protocols [6], *mte (minimum transmission energy)* lasts for the longest time. However, we observe from Fig. 5 that only one or two nodes are really alive. As such, *leach* and *leach-c* (a variant of *leach*) can be considered the most efficient protocols, in terms of both energy dissipation and longevity.

In Figures 6 and 7, we compare the two protocols. We see that both modes of TEEN perform much better than *leach*. If the cluster formation is based on the *leach-c* protocol, the performance of the TEEN protocol is expected to be correspondingly better.

As expected, *soft mode* TEEN performs much better than

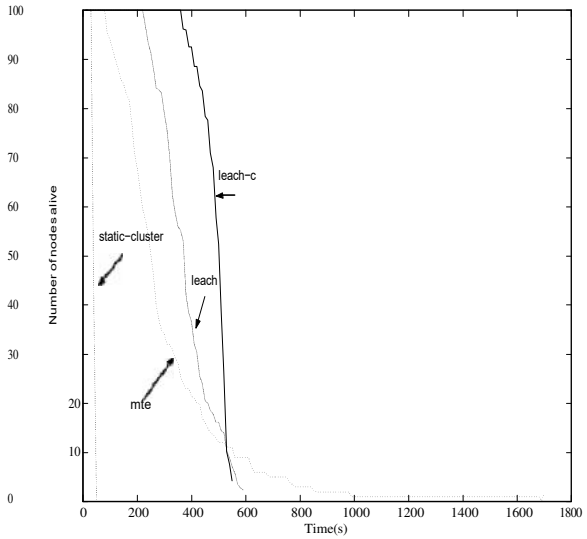


Figure 5. *No. of nodes alive: LEACH*

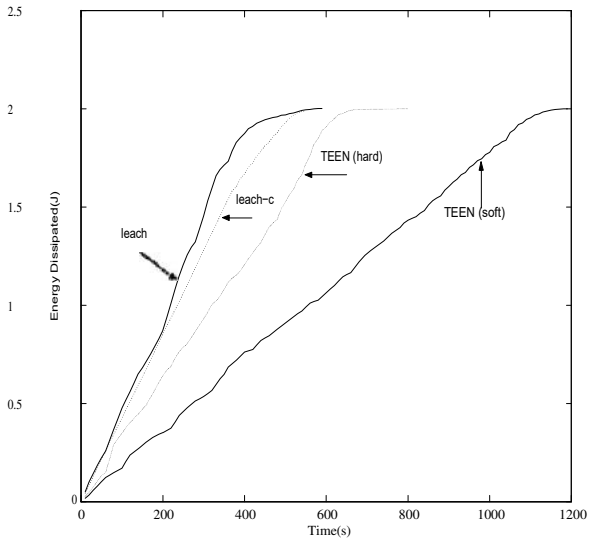


Figure 6. *Comparison of average energy dissipation*

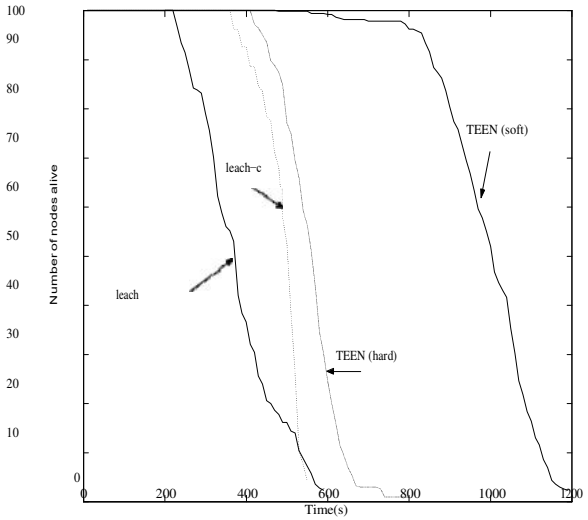


Figure 7. *Comparison of the no. of nodes alive*

hard mode TEEN because of the presence of the soft threshold.

8. Conclusions

In this paper, we present a formal classification of sensor networks. We also introduce a new network protocol, *TEEN* for reactive networks. TEEN is well suited for time critical applications and is also quite efficient in terms of energy consumption and response time. It also allows the user to control the energy consumption and accuracy to suit the application.

Acknowledgment

We would like to thank Wendi Heinzelman for her valuable suggestions and for letting us use her *LEACH* extension to the *ns* simulator for our experiments.

References

[1] J. Broch, D. Maltz, D. Johnson, Y. Hu, and J. Jetcheva. "A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols". In *Proceedings of the Fourth Annual ACM/IEEE International Conference on Mobile Computing and Networking(MOBICOM), ACM*, October 1998.

[2] J. Elson and D. Estrin. "An Address-Free Architecture for Dynamic Sensor Networks". Technical Report 00-724, Computer Science Department, USC, January 2000.

[3] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar. "Next Century Challenges: Scalable Coordination in Wireless Net- works". In *Proceedings of the 5th Annual ACM/IEEE In- ternational Conference on Mobile Computing and Network- ing(MOBICOM)*, pages 263–270, 1999.

[4] W. Heinzelman, A. Chandrakasan, and H. Balakrish- nan. *"uAMPS ns Code Extensions"*. http://www-mtl.mit.edu/research/icsystems/uamps/ leach.

[5] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan. "Energy-Efficient Communication Protocols for Wireless Microsensor Networks". In *Proceedings of Hawaiian In- ternational Conference on Systems Science*, January 2000.

[6] W. B. Heinzelman. *"Application-Specific Protocol Archi- tectures for Wireless Networks"*. PhD thesis, Massachusetts Institute of Technology, June 2000.

[7] C. Intanagonwiwat, R. Govindan, and D. Estrin. "Di- rected Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks ". In *Proceedings of the 6th Annual ACM/IEEE International Conference on Mobile Computing and Networking(MOBICOM)*, pages 56–67, Au- gust 2000.

[8] M. Jiang, J. Li, and Y. C. Tay. "Cluster Based Routing Pro- tocol". Internet Draft, 1999.

[9] E. M. Royer and C.-K. Toh. "A Review of Current Rout- ing Protocols for Ad-Hoc Mobile Wireless Networks". In *IEEE Personal Communications Magazine*, pages 46–55, April 1999.

[10] UCB/LBNL/VINT. *"Network Simulator-ns"*. http://www- mash.cs.berkeley.edu/ns.

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

Efficient Irrigation System using IOT And Raspberry Pi

Puladas Sandhya Priyanka[1] , K.John Bunyan[2]

[1] *Assistant Professor, Dept. of CSE, Malla Reddy College of Engineering, Secunderabad*
[2] *Assistant Professor, Dept. of CSE, Malla Reddy College of Engineering, Secunderabad*

Abstract

Water is the important source in human life. Around 80 % to 90 % water used in agriculture field. As due to day by day growth in globalization and population water consumption is also increases. There is challenge in front of every country to reduce the farm water consumption and provide fresh and healthy food. Today automation is one of the important role in human life. The system is not only provides comfort but also reduce energy, efficiency and time saving. Whenever there is a change in temperature, humidity and current status of rain of the surroundings these sensors senses the change in temperature and humidity and gives an interrupt signal to the raspberry pi. Now a day the industries are using an automation and control machines which are high in cost and not suitable for using in a farm & garden field. So in this work we design a smart irrigation technology based on IOT using Raspberry pi. The system can be used to control the water motor automatically and can also monitor the growth of plant by using webcam. We can watch live streaming of farm on mobile phone using suitable application by using Wi-Fi network. Raspberry pi is the main heart of the overall system.

Key Words*: Raspberry Pi, Wi-Fi, Sensors, IOT, automation*

I. INTRODUCTION

India is one of the largest freshwater user in the world, and our country uses large amount of fresh water than other country. There is a large amount of water used in agriculture field rather than domestic and industrial sector. 65% of total water is contributes as a groundwater. Today water has become one of the important source on the earth and most of used in the agriculture field. As the soil-moisture sensor and temperature sensor are placed in the root zone of the plants, the system can distributed this information through the wireless network. The raspberry pi is the heart of the system and the webcam is interfaced with Raspberry pi via Wi-Fi Module. Python programming language is used for automation purpose. The system is a network of wireless sensors and a wireless base station which can be used to provide the sensors data to automate the irrigation system. The system can used the sensors such as soil moisture sensor and soil temperature sensor and also ultrasonic sensor. The raspberry pi model is programmed such that if the either soil moisture or temperature parameters cross a predefined threshold level, the irrigation system is automated, i.e. the relay connected to the raspberry pi will turn ON or OFF the motor. This paper present an efficient, fairly cheap and easy automated irrigation system. This system once installed it has less maintenance cost and is easy to use. By using the webcam with suitable application on mobile phone we can easily online monitoring the actual situation of the field and sensors such as soil moisture and temperature are used to provide the information about changes occurs in the field. It is more advantageous than the traditional agriculture techniques.

II. RELATED WORK

After extensive research in the agricultural field, many researchers found that the agriculture area and its productivity are decreasing by the day. With the Use of different technology in the field of agriculture we can increase the production as well as reduce manual efforts. This paper shows the technology used in agriculture sector based on IOT and Raspberry Pi. Chandan kumar Sahu proposed a system on "A Low Cost Smart Irrigation Control System". It includes a number of wireless sensors which are placed in different directions of the farm field. Each sensor is integrated with a wireless networking device and the data received by the "ATMEGA318" microcontroller which is on the "ARDUINO-UNO" development board. The Raspberry pi is used to send various types of data like text messages and images through internet communication to the microcontroller process [1]. Supraha Jadhv proposed, automated irrigation system using wireless sensor network and raspberry pi that control the activities of drip irrigation system efficiently [2]. Sebastian Hentzelt proposed a paper on the water distribution system and gave results to decompose the original nonlinear optimal control problem (OCP) [3]. Joauin Gutierrez attempted a paper that research automated irrigation system using a wireless sensor network and GPRS module instead of the Raspberry pi [4]. Ms. Deweshvree Rane Proposed "Review paper based on Automatic Irrigation System Based on RF Module" it is based on the RF module, this device is used to transmit or received radio signal between two devices. It's design is complex because of the sensitivity of radio circuits and the accuracy of the components [5]. Karan Kansara proposed "Sensor based automatic irrigation system with IoT", this irrigation system is used a rain gun pipe, one end connected to the water pump and another to the root of plant. It doesn't provide water as a natural rainfall like sprinkler and also it uses only soil moisture sensor [6]. G. Parameswaran proposed "Aurdino based smart irrigation system using Internet of Things", the researcher has not used Raspberry pi instead the work is done using aurdino controller without use of soil moisture sensors [7].

III. PROPOSED SYSTEM

The block diagram of the proposed system is as shown in the below figures. Fig a. represents the Transmitting Section whereas Fig b. represents the Receiving Section. The main components of this diagram are Sensors, Raspberry Pi module, Wi-Fi connection, LDR, relay, motor, and lamp.

ISBN 978-93-88808-61-3

Block Diagram

a. Transmitting section

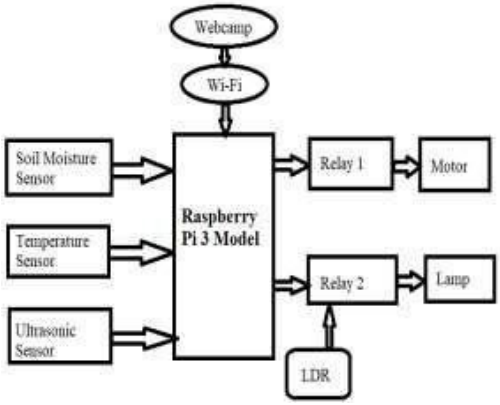


Fig a: Irrigation Control System (Transmitting Section)

The above figure shows that main block diagram of Irrigation control system. In that main model is Raspberry pi 3 model, Relays, LDR, Sensors. In this control system three sensors are such as soil moisture sensor, temperature sensor, ultrasonic sensors are connected to the raspberry pi 3 model also Wi-Fi connection is connected to the model. The connection of raspberry pi is given to the relay 1 and relay 2 which are again given to the motor and lamp respectively. LDR connection is given to the relay 2.

b. Receiving section

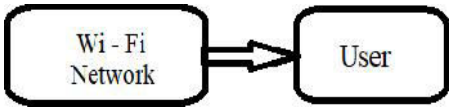


Fig b: Monitoring Unit (Receiving Section)

Above figure shows that receiving section of the main module i.e Monitoring unit. The two sections present are: one is Wi-Fi network and user. This connection again given to the raspberry pi 3 module.

SENSORS

A sensor is a device, module, or subsystem whose purpose is to detect events or changes in its environment and send the information to other electronics, frequently a computer processor. In short sensors are the device which converts the physical parameter into the electric signal. A sensor's sensitivity indicates how much the sensor's output changes when the input quantity being measured changes. The system which shown in fig.1 consists of

- Soil moisture sensor- used to measure the moisture content of the soil.
- Temperature sensor - used to detect the temperature of the soil.
- Ultrasonic sensor - used to measure the water level in the water tank.

Fig 3.2.1: Soil Moisture Sensor

Fig 3.2.2: DHT 11(Temperature Sensor)

RASPBERRY PI

Raspberry Pi is a small sized single board computer which is capable of doing the entire job that an average desktop computer does like spread sheets, Word processing, Internet, Programming, Games etc. It contain 1GB RAM, 2 USB, ARM V8 Processor and an Ethernet port, HDMI & RCA ports for display,3.5mm Audio jack, SD card slot (bootable), General purpose I/O pins, runs on 5v.

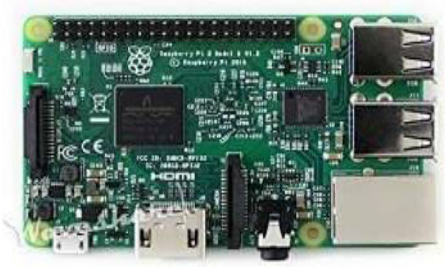


Fig 3.3: Raspberry Pi Model

RELAY

A relay is an electrically operated switch. Relays are used where it is necessary to control a circuit by a separate low-power signal. A relay with calibrated operating characteristics and sometimes multiple operating coils are used to protect electrical circuits from overload. As shown in above figure raspberry pi is connected to the devices via relay. Here relay can be operated as switch to on or off the devices.

Fig 3.4: Relay

IV. WORKING PRINCIPLE

As the Raspberry Pi is the heart of the system. This system contain webcam which is interfaced to Raspberry Pi via Wi- Fi module. The Raspberry Pi Model zero incorporates a number of enhancements and new features. This features of raspberry pi are improved power consumption, increased connectivity and greater IO which made this powerful, small and lightweight ARM based computer. The Raspberry Pi cannot directly drive the relay. It has only zero volts or 3.3 V. It needs 12V to drive electromechanical relay. In that case it uses a driver circuit which provides 12V amplitude to drive the relay. Various sensors are connected to the Raspberry Pi board give a resistance variation at the output. This output signal is applied to the comparator and signal conditioning circuit which has potentiometer to decide the moisture level above which the output of comparator goes high. This output signal is given to the Raspberry Pi board. If the soil moisture value is above the moisture level then the 3 phase induction motor will be OFF, whereas if the moisture level is low motor will be ON through the relay. LDR (Light Dependent Resistor) is used to control the light automatically and by using this we can monitor the farm at night also.

V. WORK FLOW OF THE SYSTEM

Step 1: Start.

Step 2: The system can be initialize on Raspberry Pi.

Step 3: The water level sensor constantly checks for the water level of the motor.

Step 4: The soil moisture sensor checks the soil moisture level constantly.

Step 5: The USB camera installed with the Raspberry Pi gives the complete lookout of the field and this can be monitored in the internal network system.

Step 6: The sensor constantly senses the temperature and humidity of the field and updates the date in the web server.

Step 7: If the permissible level of water is reduces, then the relay which is connected to the Raspberry Pi will turn ON the motor.

Step 8: Similarly, if the soil becomes dry, the motor which is connected to the relay will be turned ON to wet the field.

Step 9: If the step 8 is completed, it will go to the step 4.

Step 10: Similarly, if the step 7 is over, the command will go to the step 3.

VI. HARDWARE PART AND RESULT

ISBN 978-93-88808-61-3

Fig: Hardware part and result shown on android APP

VII. CONCLUSION

The smart irrigation system is suitable and cost effective for advance water resources for agricultural production. The system would provide feedback control system which will monitor and control all the activities of plant growth and irrigation system efficiently. If rain gun sensor can be added so that when it rains there won't be floods. Rain water harvesting can be done and this harvested water can be used to irrigate fields. We can also include many more water quality sensors that affect the crops.

VIII. REFERENCES

1. Chandankumar Sahu, Pramitee Behera. 2015 A Low Cost Smart Irrigation Control System IEEE Sponsored 2nd International Conference on Electronics and Communication System (ICECS).

2. Suprabha Jadhav, Shailesh Hambarde. 2013 Automated Irrigation System using Wireless Sensor Network and Raspberry pi, International Journal of science and research (IJSR).

3. Gutierrez, J. Francisco, J. Villa-Medina Nieto-Garibay, A., and Angel, P.G. 2013. Automated Irrigation System Using a Wireless Sensor Network and GPRS Module

4. Rane D, Indurkar P, and Khatri, D.M. 2015 Paper based on Automatic irrigation system on RF module In IJAICT Volume 1, Issue 9.

5. Kansara, K., Zaveri, V., Shah, S., Delwadkar, S., Jani, K. 2015 Sensor Based Automated Irrigation System with IOT In IJCSIT, Vol. 6.

6. Parameswaran, G., Sivaprasath, K. 2016 Arduino Based Smart Drip Irrigation System Using Internet Of Things In: IJESC, Volume 6 Issue No. 5.

ISBN 978-93-88808-61-3

Text Mining and Exploration using SVM

Arun.k[1] Syed Mohammed Shafi[2]
[1] Mallareddy college of Engineering, CSE dept, HYD
[2] Mallareddy college of Engineering, CSE dept,HYD,

Abstract— **On the basis of analyzing the basic concepts and the process of text excavation, the present study proposes some new methods in extraction of text features, deflation of characteristic collection, extraction of study and knowledge pattern, and appraisal of model quality. Meanwhile, it makes a comparison of two types of text categorization, text classifications and text cluster, and it briefly explores the basic issues to be solved in the future development of the text excavation technology.**

Key words- Text excavation; Text Features; characteristic collection deflation; Text classification; Text cluster

Ⅰ. INTRODUCTION

Along with the Internet application's popularization, Web already developed into has 300,000,000 page's distributional information spaces, moreover this digit still by the speed which every half year doubles grew. In the middle of these mass data, the majority of information are the non-structurization perhaps half structurization, moreover is containing the huge potential value knowledge. The people urgent need can from Web fast, discover these valuable knowledge effectively. On Web the information multiplicity has decided the Web knowledge discovery multiplicity. According to the processing object's difference, may the Web knowledge discover that divides into two broad headings: Content discovery and structure discovery. The content discovered that is mainly the excavation which keeps off to article this article. The text excavation (TextMining), may the massive documents set content carry on the abstract, classified, the cluster, the connection analysis as well as to Web on carries on the tendency forecast to the documents and so on.

Ⅱ.BASIC CONCEPTS

The text is by the massive characters, the word, the sentence is composed, to text excavation, in paramount consideration text character word. In English, Chinese and so on the natural language, have the massive words the concurrently kind of phenomenon, this for the text part-of-speech tagging, semantic labelling has brought the very major difficulty. Therefore,how to remove the part of speech, the semantic different meanings, is the text automatic labelling research key question.

A. A part-of-speech taggings

Ⅰ*)Concurrently kind of word:* Has two or two above lexical category glossary calls the concurrently kind of word.

the concurrently kind of word displays the different semantics in the different context linguistic environment, is by the concurrently kind of word lexical category decided that this is in the text excavation part-of-speech tagging question.

Concurrently kind of word classification Same-type opposite sex different righteousness concurrently kind of word

For example: Chairman Mao leads us to fight for state power. (“leadership” is a verb, leads, meaning of the instruction)

Chairman Mao is our good leadership. (“leadership” of is noun, meaning of person in charge, the leader)

Same-type opposite sex synonymy concurrently kind of word

For example: He has worked for 3 hours. (“hour” is classifier, Unit of time)

We measure the operating time by the hour. (“hour” is noun, Unit of time)

Heterogeneous homogeneous synonymy concurrently kind of word

For example: The computer has bought 50 computers. (“computer” is noun and “computer” synonymy)

The computer has bought 50 computers. (“computer” is noun and “computer” synonymy)

The non-word usage (stops word usage): In text relatively auxiliary functional word.

Ⅱ)*Non-word usage classification*

Function word: In English “a, the, for, with,…”; In Chinese “,…”And so on.

Full word: In database conference’s paper “database” a word, although the frequency of use is very high, but regards as the non-word usage.

Ⅲ *) Stem question:*compute, computes, computed identifies a word (distortion).

Ⅳ*) part-of-speech tagging:* The so-called part-of-speech tagging is for the text in word labelling part

of speech. Is mainly refers to the concurrently kind of word the lexical category to determine that the concurrently kind of word's lexical category determines only the sentence in according to the context.

B semantic labelling

semantics labelled a word to be equivocal, has formed the word different meanings phenomenon, semantic labelling mainly solves the word different meanings problem. A word equivocal is also in the natural language common phenomenon, but, in certain context, a word can only explain that generally is one semantics.

semantics labelling is to appears the words and expressions semantics carries on the determination in certain context, determined that its correct semantics and labels.

Semantic automatic labelling method Usual word is composed of meaning

The related word's method which appears using the retrieval context in determines the polysemant righteousness item

Determines the polysemant using the context matching relations the word meaning

To dispel equivocally with the most greatly possible righteousness item

C labelling technologies

The commonly used labelling technology route is based on the probability statistics and based on the rule method.

I *)Based on probability statistics CLAWS algorithm*

CLAWS is English Constituent-Likelihood Automatic Word-tagging System (ingredient likelihood automatic lexical category automatic labelling system) one algorithm which the abbreviation, it was in 1983 Ma Shaer (Mashall) when gives the LOB corpus (to have each literary style British English corpus, storage capacity quantity is 1,000,000 words) made automatic part-of-speech tagging proposed

II*) Based on probability statistics VOLSUNGA algorithm*

The VOLSUNGA algorithm is to the CLAWS algorithm improvement, in optimal path's choice aspect, is not only then calculates the probability to accumulate the biggest mark string finally, but along direction from left to right, the use “fortifies at every step” the strategy, regarding the current consideration's word, only retains leads to this word the optimal path, discards other ways, then embarks again from this word, carries on the match this way with next word's all marks, continues to discover the best way, discards other ways, goes forward like this gradually, walks until the entire cross section, obtains the entire cross section the optimal path to take the result output. Counts each word according to the corpus the relative labelling probability (Relative Tag Probability), and is auxiliary the optimal path with this kind of relative labelling probability the choice. The VOLSUNGA algorithm reduced the CLAWS algorithm time order of complexity and the spatial order of complexity greatly, raised the automatic part-of-speech tagging rate of accuracy.

The CLAWS algorithm and the VOLSUNGA algorithm are based on the statistical automatic labelling method, acts according to merely with the present probability labels the lexical category. But, with the present probability is only the biggest possibility and is not the only possibility, by determines the concurrently kind of word with the present probability, is by discards with the present probability low possible premise. In order to enhance the automatic part-of-speech tagging the accuracy, but must auxiliary by based on the rule method, determines the concurrently kind of word according to the language rule.

D other text retrieval labelling technology

I *) Inverted index*

Inverted index is an index structure that contains two hash tables index table, or two B +-tree index table, shown in Table 1, Table 2.

Table 1 Document Table (document_table)

doc_ID	posting_list
Doc_1	t_1_1, ..., t_1_n
Doc_2	t_2_1, ..., t_2_n
⋮	⋮
Doc_n	t_n_1, ..., t_n_n

Table 2 vocabulary (term_table)

term_ID	posting_list
Term_1	doc_1, ..., doc_i
Term_2	doc_1, ..., doc_j
⋮	⋮
Term_n	doc_1, ..., doc_n

Table 1 is composed of a group of documents record, posting_list is appears in the documents the word tabulation; Table 2 are composed of a group of word record, posting_list is contains this word the documents marking tabulation. Through such two tables, may discover with and assigns the documents related all words as well as with group of word related all documents for the decisive remark collection related all documents. But cannot process the synonym and the polysemant question, and posting_list is long, causes the memory expenses to increase.

II*) Signature File*

Features file is a storage database, the characteristics of each record of a document file. A feature of each bit corresponds to a fixed-length string, a bit corresponds to a word, if a particular word corresponds to appear in the document is, then the location of one, otherwise set 0.

IIITEXT EXCAVATION PROCESS

The text excavation object usually is group of HTML perhaps the XML form documents collection. Text excavation's general treating processes is: Document Set,Characteristics of the establishment of,Reduced feature set, Learning and knowledge extraction model,

Model Quality Evaluation-Knowledge model.

ISBN 978-93-88808-61-3

A . Text Features

Text feature refers to the metadata on the text. It can be divided into descriptive features (text, name, date, size, type, etc.) and semantic features (text, author, title, organization, content, etc.). Text feature to feature vectors, said: , Where t_i for the entry entry, $w_i(d)$ for t_i in d in the weights. As the feature vector dimension is usually very high, generally use the evaluation function to carryout feature selection. Evaluation of commonly used functions: nformation Gain,Expected Cross Entropy, Mutual Information,the Weight of Evidence for Text, Word Frequency.

Document Modeling：Using vector space model (VSM) of the text document model.

Frequency Matrix: line corresponds to the word *w,* the column vector corresponding to the document *d,* the simplest vector of values of words in the document appears on the value of 1, otherwise value is 0, Table 3 is based on occurrences of the word for the word frequency vector matrix, the value to reflect the word *w* and a document *d* of the correlation.

Table 3 Frequency of the Frequency Matrix document

	d_1	d_2	d_3	d_4	d_5	d_6
w_1	322	85	35	69	15	320
w_2	361	90	76	57	13	370
w_3	25	33	160	48	221	26
w_4	30	140	70	201	16	35

With the similarity of the document word frequency matrix can be measured, the typical method is the cosine similarity metric calculation (Cosine Measure).

Cosine similarity definition: $SIM(v_1, v_2) = \frac{v_1 \cdot v_2}{|v_1||v_2|}$, is two documents vectors, The inner product is the standard vector dot product, Defines for $\sum_{i=1}^{w} v_{1i} v_{2i}$, is defined for $\sqrt{v_1 \cdot v_1}$, $|v_1|$ is direction

B characteristic collection deflation

Term frequency matrix similarly Gao Weishu, sparse data influence, to overcome these questions, the people proposed the latent semantic index (Latent Semantic Indexing) the method reduces the characteristic collection.

I) *Latent semantic index* ： "The singular value decomposes (Singular Value Decomposition using the matrix theory, SVD)" the technology, transforms the term frequency matrix as the singular matrix (K×K), concrete step:

- Establishment term frequency matrix, frequency matrix.

  Calculates frequency matrix the singular value decomposition. Decomposes frequency matrix to become 3 matrix U, S, V. U and V is the orthogonal matrix (UTU=I), S is the singular value diagonal matrix (K×K).

  Regarding each documents d, after removing in SVD eliminates the word new vector replace original vector.

  Preserves all vector set, founds the index with the high-level multi-dimensional index technology for it.

  Carries on the similarity computation after the transformation documents vector.

C studies and knowledge pattern extractions

I) *Participle*： The participle refers to between the text word and the word adds on the blank space, refers to Chinese text, because between English itself word and the word is differentiates by the blank space.

II) *Automatic participle*： The automatic participle is refers to uses the computer adds on the blank space automatically between the word and the word. The use is:

a) Chinese text automatic retrieval, filtration.
b) Classification and abstract.
c) Chinese text automatic proofreading.
d) Outside Chinese machine translation.
e) Chinese character recognition.
f) Chinese speech synthesis.
g)Take sentence as unit's Chinese character keyboard entry.
h)Chinese character Jan traditional form transformation.

III) *Main participle method*

a) Biggest match law (Maximum Matching method, MM law): The selection contains 6-8 Chinese characters the strings to take the biggest string, matches in the biggest string and the dictionary word clause, if cannot match, slices off a Chinese character to continue to match, until found the corresponding word in the dictionary. The match direction is from right to left.

b) Reversion biggest match law (Reverse Maximum method, RMM law): The match direction and the MM law are opposite, is from left to right. The experiment indicated: Regarding Chinese, the reversion biggest match law is more effective than the biggest match law.

c) Bilateral matching law (Bi-direction Matching method, BM law): Compared with the MM law and RMM law participle result, thus decides the correct participle.

d) Optimum matching law (Optimum Matching method, OM law): The dictionary in word according to them in the text appearance frequency's size arrangement, the high frequency's word arranges before, the frequency low word arranges, thus enhancement match speed.

e)Association backtracking: Uses the mechanism which associates and recalls to carry on the match.

IV)*Feature extraction*

ISBN 978-93-88808-61-3

The feature extraction is the glossary which, the phrase feature extraction appears to the text.

Characteristic weight function:

$$f_w(t_i) = \frac{f_v(t_i)\log(1+f_v(t_i))}{\sqrt{\sum_{j=1}^{m}(f_v(t_j)\log(1+f_v(t_j)))^2}}$$

And: Expresses the characteristic weight function; Expresses the characteristic item in the text frequency; Expressed that the characteristic paragraph frequency, namely contains ti the paragraph number/text total paragraph number.

V)*Automatic digest*

The automatic digest is uses the computer to withdraw automatically from the primitive documents reflects this documents center content accurately comprehensively the simple coherent short written work. Our country in 1995 carried on to the automatic digest system has evaluated, the system which for the first time participated has 3. The evaluation result performance is:

1. Three systems may according to the ratio which assigns from the original text select part of sentences.
2. The extraction digest is in the original text sentence, only then in the system 2 digests has rejected some digit.
3. Three system's digests do not superpose nearly completely.

may see the automatic digest system from above result also to have many foundation work to do. the text abstract is refers to from the documents extracts the key information, carries on the abstract or the explanation with the succinct form to the documents.

Thus, the user does not need to glance over the full text to be possible to understand the documents or the documents set overall content. The text abstract is very useful in some situations, for example, search engine when to user returns inquiry result, usually needs to give the documents the abstract.

D. model quality appraisal

Carries on the excavation in the text to be possible to regard as is one kind of machine learning process. The study result is the knowledge model, carries on the appraisal to the knowledge model is the machine learning important component. The typical assessment method is to the text retrieval basic measure.

{relevant}: With some inquiry related documents set.

{retrieved}: The system retrieves documents set.

{relevant} ∩ {retrieved}: Both are related and the actual documents set which retrieves.

precision: Both are related and the actual documents which retrieves with the documents percentage which retrieves.

recall: Both are related and actual documents which and the inquiry related documents percentage retrieves.

IV. TEXT CLASSIFICATIONS

The text classification is refers to according to the subject category which defines in advance, determines a category for documents set's in each documents. Thus, not only the user can glance over the documents conveniently, moreover may make the documents through the limit hunting zone the search to be easier. At present some websites use the man-power to carry on the classification to the Web documents, some websites use the automatic sorting. The text classification technology algorithm has many kinds, the commonly used algorithm has TFIDF and Nave Bayes and so on.

A Generally method

Will have classified in advance the documents take the training regulations.

Obtains the disaggregated model from the training regulations (to need test procedure, unceasing refinement).

With the disaggregated model which derives to other documents classifies.

B Based on connection taxonomic approach

Proposes the key words and the glossary through the information retrieval technology and the connection parsing technique.

Uses the existing part of speech production key words and the word concept level (documents category).

Discovers the associated word using the connection excavation method, then differentiates each kind of documents (each kind of documents to correspond a group of connection rule).

Goes with the connection rule to the new documents classification.

C Web documents automatic sorting

Uses in the ultra link the information to carry on the classification, the commonly used method includes:

Statistical method

Markov random field (Markov Random Field, MRF)

Unifies loose marking (Relaxation Labeling, RL)

V. TEXT CLUSTERS

The text cluster and the classified difference lies, the cluster has not defined the good subject category in advance, its goal is divides into the documents certain kinds, the request identical kind in documents content similarity is as far as possible big, but the different kind of between similarity is as far as possible small. Hearst et al. the research had already proven “the cluster supposition” the question, namely approaches with the inquiry related documents cluster's comparison, and is far away from the non-correlated documents. Therefore, the documents which will search using the cluster technology divides into certain

kinds. At present has many kinds of text cluster algorithm. Divides into two big types approximately: Level cluster and plane allocation method.

A Level cluster law

Concrete process:

Documents collection D= {d1,…, di,…, dn} each documents di regards as has single member's kind of ci={di}, these kinds constituted D cluster C= {c1,…, ci,…, cn};

Calculates in C every time to the kind (ci, cj) between similarity SIM (ci, cj);

The selection has the biggest similarity kind to arg max SIM (ci, cj), and ci and the cj merge is one new kind ck=ci ∪ cj, thus constitutes D new kind of C= {c1,…, cn-1};

Is redundant the above step, is only left over one kind until C.

Materially this process constructed one to contain in the kind of level information as well as during all kinds and the kind of similarity spanning tree. Because each time merges time, needs overall situation quite all kind of between the similarity, then choice best two kinds, therefore the operating efficiency is not high, does not suit in the massive documents set.

B Plane allocation method

The plane allocation method is documents collection D= {d1,…, di,…, dn} horizontal divides for certain kinds, concrete process:

The determination must produce kind of numberk;

Produces k cluster center according to some kind of principle to take the cluster seed S= {s1,…, sj,…, sk};

To D each documents di, calculates it and each seed sj similarity SIM in turn (di, sj);

The selection has biggest similarity seed arg max SIM (di, sj), belongs to di take sj as cluster center kind of Cj, thus obtains D cluster C= {c1,…,ck};

The redundant step 2~4 certain times, by obtains the stabler cluster result.

This method speed is quick, but k must determine in advance, seed selection difficulty.

the text cluster also has the k-means algorithm, the simple Baye cluster law, the K- most close neighbor to refer to the cluster law, the graduation cluster law as well as based on the concept text cluster and so on.

VI RELATED CONTENTS

Text excavation besides above several introduction content, but also has the following related content research:

Chinese character input and Chinese corpus.

Text phrase delimitation and syntax labelling.

Electronic dictionary construction.

Terminology database.

Machine translation.

Computer auxiliary text proofreading.

Information automatic retrieval system.

Chinese speech recognition system.

Chinese speech synthesis system.

Chinese character recognition system.

The related text excavation's product model has the IBM text intelligence excavator (the hard core is TextMiner, its major function is the feature extraction, the documents accumulation, the documents classification and the retrieval; Supports 16 languages many kinds of form text data retrieval; Uses the deep level the text analysis and the index method; Supports the full-text search and the index search, the search condition may be the natural language and the Boolean logical condition.), the Autonomy Corporation most core's product is Concept Agents (can extract concept automatically from text) as well as Tsinghua University's TH-OCR Chinese character recognition system (recognition precision reaches above 98%).

VIICONCLLUSIONS AND FORECAST

The text excavation, needs to use the natural language processing technology inevitably, constructs the large-scale real text the corpus is the most foundation work. This article elaborated the content is in the text excavation key job. if the foundation work is not solid, the text excavation is very difficult on a big stair. Basic research's foresightedness ought to be able to guarantee in technical the sophistication. Future text excavation technology should be the knowledge retrieval, the knowledge retrieval development should be able the effective addressing following some key questions: (a). Structurized data and non-structurized data mix retrieval; (b) Half structurized content retrieval XML content retrieval; (c).Engine intellectualization knowledge retrieval.

REFERENCE

[1] C.Faloutsos. *Access Methods for Text.* ACM Comput. Surv. , 17 p49-74, 1985.

[2] G.Salton. *Automatic Text Processing*. Reading, MA: Addison-Wesley,1989.

[3] C.J.Van Rijsbergen. *Information Retrieval.* Butterworth,1990.

[4] C.T.Yu and W.Meng. *Principles of Database Query Processing for Advanced Applications*. San Francisco: Morgan Kaufmann,1997.

[5] K.Wang,S.Zhou,and S.C.liew. *Building Hierarchical Classifiers Using Class Proximity.* In Proc.1999 Int. Conf. VLDB'99,P363-374,Edinburgh,UK,Sept.1999.

[6] P.Raghavan. *Information Retrieval Algorithms:A Survey.* In proc.1997 ACM-SIAM Symp. Discrete Algorithms, p11-18,New Orleans,LA,1997.

[7] J.M.Kleinberg and A.Tomkins. *Application of Linear Algebra in Information Retrieval and Hypertext Analysis.* In proc. 18th ACM Symp. Principles of Database Systems,P185-193,Philadelpgia,PA,May 1999

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

A Noval System For Early Detection Of Thyroid With Graph Cluster Ant Colony Optimization

Sayyad Rasheeduddin,
Asst.Professor,
CSE Department,
Malla Reddy College of Engineering

Ch.Vijaya Kumari
Assoc. Professor,
CSE Department,
Malla Reddy College of Engineering

ABSTRACT: Thyroid nodule is defined as an endocrine malignancy that occurs in humans due to abnormal growth of cells. Recently, an increasing level of thyroid incidence has been identified worldwide. Thus, it is necessary to detect the nodules at an early stage. Ultrasonography is an important tool that is utilized for the detection as well as differentiation of malignant thyroid nodules from benign nodules.. Further, large number of features available in US characteristics increases the computation time as well as complexity of classification. In this paper, Graph-Clustering Ant Colony Optimization based Extreme Learning Machine approach is proposed to achieve efficient diagnosis of thyroid nodules. It will enhance thyroid nodule classification by selecting only the optimal features and further using it for improving the function of classifier. The main goal of this technique is to differentiate the malignant nodules from the benign nodules. The performance of both feature selection and classification are evaluated through parameters such as accuracy, AUC, sensitivity and specificity. From the experimental results, it is revealed that the proposed method is significantly better than the existing methods. Thus, it is considered to be an effective tool for diagnosing the thyroid nodules with less complexity and reduced computation time.

Keywords.: Thyroid nodule, ultrasound image, diagnosis, feature extraction, nodules classification.

I. INTRODUCTION

Thyroid is a butterfly shaped small gland situated in the lower region of neck under the layers of skin and muscle. The abnormal growth of cells in the thyroid glands referred to as thyroid nodules. These nodules may be either benign or malignant commonly called as non-cancerous or cancerous cells respectively (Acharya et al, 2016). Thyroid nodules are the most common search criteria in thyroid gland as it is present in almost 40% of the population among world-wide and about 5-10% is found to be malignant. Thus, radiologists are involved in diagnosing the thyroid gland to identify the risk of malignancy with respect to the guidelines provided by Thyroid Imaging, Reporting and Data System (TI-RADS). In general, thyroid nodules affect both men and women whereas it is severe in case of women and its formation depends upon different characteristics like gender, age and population (Erdem et al, 2010). The thyroid disease analytics have revealed that thyroid is a severe disorder which increases the mortality rate in humans.

Therefore, it is necessary to produce an accurate tool for malignancy risk detection in order to increase the survival rate of thyroid patients. Moreover, early identification of the symptoms of thyroid disorders can improve the survival rate thereby initiating the treatment at initial stage (Koundal et al, 2018). However, the diagnosing process as well as treatment of thyroid disease remains difficult and the main challenge in this field is differentiation between the nodules. It is necessary to accurately classify the thyroid nodule because of high prevalence of the nodules as well as less prevalence of the malignancy.

Fine needle aspiration biopsy (FNAB) is the standard treatment utilized for diagnosing the thyroid diseases but it is reported that it can mimic other kind of diseases (Bakshi et al, 2003).Several thyroid treatment plans use FNABs as reference since it is labor-dependent and expensive under large scale diagnosis. Likewise, unwanted biopsies cause anxiety, irritation and increase the treatment-expense to thyroid patients (Ma et al, 2017). Even though massive growth is achieved in the field of thyroid diagnostics with sources such as CT imaging, radionuclide and MRIs still it is necessary to select an appropriate and stable material for effective differentiation between the nodules (Wu et al, 2013). Furthermore, clinical procedures do not obtain better diagnosis and so, non-invasive imagery tools like Ultrasonography is identified as a best choice for distinguishing among the nodules. The American Thyroid Association (ATA) stated that ultrasound images are the primary choice of any radiologists for examining the thyroid nodules. Furthermore, if a nodule is identified on other kind of image modalities, detailed diagnostics are performed on US images. Thus, Ultrasonography is the initial stage diagnosing modality for thyroid disease identification. (Cooper et al, 2006) defined that US images are sensitive and suitable for examining the nodularity of thyroid compared to other images such as MRI and CT. Sonography visualize the different characteristics of thyroid glands like dimension, structure, echogenicity, availability of calcification, etc. In literature, more number of research works has been carried out to distinguish between benign and malignant nodules as it is necessary to provide proper and effective treatment to thyroid patients. Thyroid nodules are comprised of different kind of textural features. Ultrasound images also resemble numerous features like electrographic, textural and morphologic which are important for the purpose of nodule classification. In the last decade, medical assistants experimented different sonographic features to prove its efficiency in diagnosing the risk of malignancy of thyroid disorders. However, feature selection is the main task of many machine learning based disease classification techniques Feature Selection 2010). A number of approaches have been developed to extract relevant features from US images. It is identified that the process of feature extraction and using them to train a classifier consumes more amount of time. Therefore, certain features are neglected during diagnosis inorder to reduce time and to improve the reliability of classification. A

proper methodis required to select appropriate features and to neglect irrelevant ones. Classification and prediction achieves accurate result with limited features than processing with all the available features. Thus, proper feature extraction and classification techniques are needed to attain better results in disease prediction. (Tsantis et al, 2009) presented CAD based diagnosis system that used morphologic and wavelet-based features for classifying the thyroid nodules in US images. These features are extracted based on malignancy related characteristics like calcification, irregular shape, uniformity, echogenicity, etc. The efficiency of using extracted features in classification is evaluated through two different pattern recognition algorithms such as probabilistic neural network and support vector machine. It showed that the extracted features can improve the accuracy of classifier and lowers the faults in thyroid disease identification.

In general, computation time and prediction accuracy are the important aspects that are taken into consideration during thyroid nodules differentiation. Therefore, in this research, an improved fast learning based pattern recognition tool called as ELM is utilized for the prediction of thyroid disease with US characteristics. Extreme Learning Machine (ELM) is a new learning based approach that supports single hidden layer feed-forward neural networks (SLFNs). Compared to gradient-based methodologies that iteratively adjust the parameters of neural network, ELM randomly selects the input weights as well as hidden biases for the determination of output weights by adopting the generalized inverse of Moore–Penrose (MP) analytical method. Further, it learns faster with highly generalized performance and also keeps the parameter tuning-free. Due to these properties, ELMs are widely used on classification areas like predicting patient outcomes (Liu et al, 2011), sales forecasting (Chen & Ou, 2011) and so on. Moreover, ELM proved its reliability on number of disease diagnosing tasks over other learning based classification algorithms.

In this paper, a machine learning model is proposed to achieve efficient classification of the thyroid nodules. Existing classification techniques have recommended that using an optimal feature selection process will enhance the accuracy of classifier used for nodule differentiation task. Hence, anovel feature selection method called as graph-clustering based ant colony optimization is adopted in this proposed work. It will select the discriminant features thereby making it effective for classifying the nodules within limited time.Further, Extreme learning machine based classifier differentiates the benign and malignant nodules. The maincontribution of this paper is defined as follows:

- A hybrid methodology is proposed to enhance the diagnostics of thyroid disease using ultrasound characteristics
- An optimal feature selection approach called as graph clustering based ant colony optimization tool is applied to extract the relevant features from the raw dataset
- To increase the efficiency as well as accuracy in differentiation of thyroid nodules, a computer aided diagnosis system based on extreme learning machine is also proposed. The remainder of this paper is organized as follows. Section 2 provides a literature review on different feature selection and machine learning based classification approaches. Section 3 presents a background review on graph clustering based ant colony optimization and extreme learning machine classifier. The detailed implementation of proposed methodology is given under Section 4. The experimental results and discussion of the proposed method is visualized in Section 5. Finally, the conclusion of this paper is defined in Section 6.

II. LITERATURE REVIEW

Recently, a number of thyroid disease diagnostic systems were introduced to analyze the severity of thyroid disorders using ultrasound characteristics. It includes computer aided diagnosis (CAD) (Sollini et al, 2018), deep convolutional neural networks (Li et al, 2019; Li et al, 2019), machine learning and so on. As CADs works on the principle of machine learning algorithms, it is mostly preferred by radiologists for identifying the risk of malignancy in thyroid glands. It is identified that machine learning based thyroid disease diagnostic systems would increase the accuracy of analysis with ultrasound imaging.(Ardakani et al, 2015) identified a new approach to analyze the texture of US images based on computer aided diagnosis (CAD) to distinguish the thyroid nodule as benign or malignant. The Receiver Operating characteristic Curve (ROC) analysis showed that Texture Analysis (TA) is a reliable approach which provides useful information to identify and to classify the nodules. Furthermore, this technique consumes low cost and does not need any human intervention as the entire diagnosis is performed on computers. But is tested on small datasets along with highly sensitive FNAB approach. FNABs need the help of surgical pathology to obtain more definitive results and it excludes certain data due to an indeterminate operation.

In classification based applications, features are the important factors that impact the discriminatory functioning of the classifier. Generally, it is effective to consider all the features during classification but it is redundant due to mutual correlation between them. Due to this, it is enough to select the relevant features from the available dataset which will then increase the classification accuracy of classifiers. However, the main task is to choose the suitable feature selection approach for the particular classification algorithm. On the other hand, selecting the optimal feature selection method can improve the classification accuracy but increases the time as well as computation complexity. In the last decade, number of feature selection algorithms was introduced that include random searches, heuristic, greedy and exhaustive. However, these techniques are computationally very expensive and get trapped into local optima. To overcome such situations, different kind of feature selection methods like Ant Colony Optimization (ACO) (Tabakhi et al, 2014), Genetic Algorithm (GA) (Kabir et al, 2011) and Particle Swarm Optimization (PSO) (Yong et al, 2016) are presented. Of these, ACO seems to be very affective as it is multi-agent based selection methodology. The advantages of ACO over swarm intelligence based techniques include its local and global optima ability, availability of long-term distributed storage, and utility of reinforcement based machine learning concept. (Choi et al, 2015) provided a systematic approach to quantitatively observe the features of US images on calcified

pathological thyroid nodule dataset. The features responsible for tumor malignancy were identified by means of a univariate algorithm and the nodules are differentiated by using neural network approach. The diagnostic ability of both the neural network and feature estimation algorithm sarederived from ROCs and AUCs (Area Under ROCs) respectively. However, the application of this technique is limited to descriptive 2D-calcified datasets as it is necessary to perform retrospective review on those datasets. In addition, only the dataset with surgical cases are taken for analysis and it does not consider clinically visible benign nodule dataset. Conversely, existing CADs cannot provide promising results and several radiologists have reported that its clinical usage is limited on certain practices.

(Ouyang et al, 2019) estimated the performance of linear as well as non-linear machine learning approaches in thyroid nodule malignancy identification with reference to a standard approach. The diagnostic performance analyzed through AUC showed similar AUCs on non-linear techniquescompared to linear techniques. Particularly, Kernel SVM and Random Forest algorithms attained moderately larger AUCs compared to other algorithms taken for observation. As these analyses does not perform any pre- processing or feature extraction tasks on image datasets, it is found to be an enhanced approach than the CAD system. The echographic view of thyroid nodule in thyroid imaging is referred to as texture. In mathematical model based diagnosing applications, these textures are analyzed by means of quantitative parameters. This is very helpful for CAD based disease diagnosing applications.

The heterogeneous nature of thyroid nodules, presence of different internal substances and large number of echo patterns in US images confuse the physicians and radiologists to identify the appropriate textures. Thus, textural feature extraction techniques are introduced to distinguish suitable texture patterns thereby decreasing the misdiagnosis rate. (Chang et al, 2010) tested six kinds of SVMs in using important textures and to increase the efficacy of thyroid lesion classification. The experimental outcomes proved the reliability of their method in extracting the important features from thyroid imagery. It is then compared with an existing approach called as sequential- floating-forward-selection (SFFS). This comparison showed that the performance of SVMs in feature extraction is similar to SFFS but the execution time is 3-37 times faster than SFFS.

(Shankar et al, 2018)established a kernel-based classification model to classify thethyroid nodules after selecting appropriate features from thyroid dataset. Grey wolf optimization based feature selection algorithm is adopted to improve the dataset classification. Their technique showed improved performance on dataset classification but it consumes large amount of time. (Han et al, 2006) discovered an ELM model to predict how long a non-small cell lung cancer postoperative patient can survive. This method showed accurate results in prediction and the convergence rate is faster than the ANN framework. (Zhang et al, 2007) determined the functioning of ELM model on multi-categorical classification of microarray dataset of cancer patients. It was observed that the classification accuracy, training time and the computation complexity of ELM are better than ANN and SVM classifiers. (Helmy&Rasheed, 2009) utilized ELMs **Extreme** to diagnose five types of diseases, and the classification accuracy and computational complexity for this observationis effective with reduced training dataset. (Gomathi & Thangaraj, 2010) suggested computer aided ELM lung cancer diagnostic system. The experimental outcome of this system is compared with SVM approach where ELM produce more accurate results in classification task.

(Li et al, 2012) recommended computer operated diagnostic system that works on the basis of principal component analysis and extreme learning machine. It performed thyroid diagnosis by taking into account different characteristics of ELM like simplicity, less complexity, generalized behavior, faster learning capability and less time for computation. Further, feature extraction is performed through PCA which ignores irrelevant data and consider only the appropriate features for ELM classification. This technique is efficient in classifying three different forms of thyroid disorders like hyperthyroid, hypothyroid and normal thyroid. PCA-ELM classifier is precise and it provides accurate categorization of thyroid malignancy. However, still is a challenging task to provide timely efficient feature selection approach for selecting appropriate features and using it for differentiating the nodules. Further, existing classifiers are highly complex which performs larger calculation throughout the entire operation. To overcome these issues, a hybrid approach is introduced in this paper by combining graph-clustering ant colony optimized feature extraction with extreme learning machine classifier.

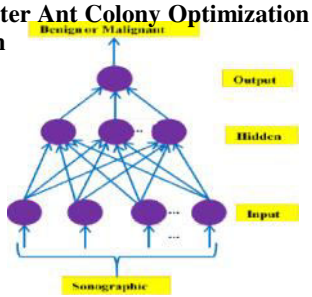
III. BACKGROUND METHODOLOGY

A *Ant Colony Optimization*

(Sivagaminathan & Ramakrishnan, 2007) explained that ant colony optimization is similar to the real-life behaviour of ants while travelling through the same path to reach their nest after collecting the food. It should be noted that it does not make any visual contact with the path itis travelling. This could be achieved through an indirect mode of communication known as “stigmery” by an odorous chemical substance named as pheromone. The quantity of the pheromone substance depends upon different properties of the food source like quantity, quality and distance of availability. All the ants follow a path that contains more pheromone thereby making it a positive feedback loop. The pheromone starts vanishing and evaporating after certain time period, and finally result in reduced pheromone in less demand path. Because of pheromone evaporation, the ants search other available paths and finalize the most optimal path for travelling. By following the same procedure, the optimal features are selected from the thyroid dataset using ant colony optimization algorithm. Due to its simplicity, it is found to be a better method for machine learning based classification designs.Nevertheless, the use of present ACO algorithms in recent application suffer from several drawbacks like computation complexity due to fully connected graph structure drawn from all the available features, need for a learning model to create feature subsets, highly correlated

ISBN 978-93-88808-61-3

with Graph Cluster Ant Colony Optimization Based Feature Selection



To address the above issues, (Moradi&Rostami, 2015) introduced a new feature selection approach by integrating ant colony optimization with graph-clustering technique. Graph Clustering based Ant Colony Optimization is a filter based multivariate feature selection process that represent the features in the form of undirected graph with nodes and edges of the graph as features and similarities between features respectively. Similar to data clustering algorithms like Fuzzy C Means (FCM) and k-means, graph clustering initially identifies the similarity among pair of data points and form an undirected graph. This graph is then divided into clusters according to the linear/non-linear boundaries estimated through an optimal objective function. In GCACO algorithm, feature clustering is performed by means of a

community detection algorithm, which produces a subset of features with each feature containing minimal redundancy with other features available in the subset. The key role of this technique over other techniques such as F-Score, L- Score, ReliefF and UFSACO is that it considers both redundancy and relevance analyses while performing the feature selection task. Likewise, GCACO identify the relationship between features before performing selection whereas univariate approaches rank the features without considering the dependency between each feature. Since graph clustering finds advantages by integrating with ant colony optimization, it is extensively used for feature selection in many applications.Graph Clustering based Ant Colony Optimization feature selection approach is adopted in the proposed work for selecting necessary features from the thyroid dataset. This is an effective approach to select discriminant features that are very essential to differentiate the thyroid nodules from the dataset.

B. Extreme Learning Machine

The use of extreme learning machine (ELM) as novel machine learning algorithm for single layer feed forward neural networks (SLFNs) displayed in Fig. 1 was first initiated by Huang et al (2004). It overcomes the drawbacks of conventional SLFNs related to slow learning speed, tuning of trivial components and improper generalization ability. Therefore, ELM possesses different properties such as fast learning capability, highly generalized performance and free parameter tuning.ELM is designed in such a manner to function well with enhanced generalization capability for performing better classification and Fig 1. Structure of Extreme Learning Machine

IV. PROPOSED METHOD

The proposed method is designed to predict the thyroid disease from dataset by classifying the thyroid nodules using US features. Firstly, the discriminate features are partitioned from the dataset using Graph Clustering based Ant Colony Optimization feature selection method. Secondly, each of the selected features is experimented to differentiate the type of nodule using Extreme Learning Machine algorithm.

A.Feature Extraction using Graph Cluster Ant Colony Optimization

GCACO is a multivariate feature selection strategy that selects the optimal features by performing dependency analysis on features structured as an undirected graph**.** In order to select the discriminant features, the relevance analysis is performed by means of the Fisher score (F-Score) and the multiple discriminant analysis (MDA).

i. Graph formation. Initially, the features are analyzed one- by-one to know about its redundancy. This analysis is achieved by creating a weighted undirected graph with all the available features. The graph representation is defined as: , where depicts the nodes of the graph, represents the edges between the graph nodes, and denotes the weight of edges of the graph. The weight between nodes and is calculated as: regression. On comparing the learning processes of both ELM as well as SLFNs gradient based iterative and back propagation methods, ELM learns faster than SLFNs.

Where, and are the features and and are the mean value of the feature vectors. The feature vectors and is either extremely correlated or uncorrelated when the weight of their edges produce a value one or zero respectively. The weight value in GCACO is normalized by means of softmax scaling technique inorder to get rid of the impact of outliers. The procedure of softmax scaling is defined as follows:

Where and are the mean and standard deviation of all values of .

A. Feature clustering. Redundancy analysis in GCACO is performed by an effective algorithm called as Louvain community detection. In this, the weighted undirected graph is partitioned into communities or sub-nodes depending upon the similarities between features (highly correlated). During initialization, each node (feature) is treated as an individual community. For each iteration, two neighbors (say and) of a node is chosen. A modular gain factor is evaluated by eliminating the node from its own community and inserting it to one of the communities of node . This procedure is continued until all the Discriminant Analysis (MDA) and the final subset of features is sorted according to their pheromone value. The implementation process of GCACO algorithm is explained below:

Step1:The algorithm is initialized by setting up the following parameters: total iterations , number of ants , evaporation coefficient of pheromone , initial pheromone quantity and other constants ,□, .

Step2: In GCACO, the relevant feature analysis isdone by the Fisher score value. Based on the F score value, the features are sorted according to their significance. The Fisher score for the th feature in the feature set is defined as:

neighbors of node are visited. The node is then added into the community that results in higher positive modular ain factor. If the modular gain of all the neighboring communities is negative, then remain in its own community. This procedure is repeated until no change is found in the modular gain value and a new network is drawn based upon the final communities. The modularity gain factor that is obtained after inserting the isolated node to any of one of the communities is determined as: Where and are mean and standard deviation of the th class with samples respectively, and is the mean of the samples in the th feature vector. The F score value is normalized within range 0 and 1 by the softmax scaling method. The features ith largest F-score value are considered to be better discriminate features.

Step 3: In GCACO, the redundancy analysis is achieved by calculating the absolute value of Pearson's correlation. To obtain this, the cross-correlation mean values within th feature and all other features visited by the th ant from all previous clusters is first evaluated. Finally, the following function is estimated to know about the redundancy:

Where is the total weight of interior edges of community , is the total weight of edges that are incident to nodes in community , is the total weight of edges that are connected to node is the total weight of edges from node to every node in community, andThis function is utilized to know about exploitation/exploration of feature during feature selection task. In this function, represents the size of .To enhance the outcomes of feature , then the th ant choose the succeeding feature as

clustering, the weights less than , a preset threshold value used to control the amount of clusters are excluded, where ranges between 0.3 and 0.8.

ii.Ant Colony Optimization. The working of Ant Colony Optimization algorithmis based on the movement of ants through the path travelled by other ants identified by pheromone chemical imprintsleft by them.In GCACO, anant is randomly allocated to one of the clusters produced by feature clustering approach. On each iteration, the algorithm considers two random values and along with a parameter ( ) and threshold value (□). The value of , , and □ lies between therange 0 and 1. If , the ant selects a feature from the cluster on the basis of roulette wheel concept. If □, the ant remains in the same cluster follows:

Where represents the features that are still not visited by the th ant from the present cluster (th cluster), is the total pheromone of the th feature, and and are the relative importance of the pheromone value and heuristic information, respectively.

(ii) If , a probability function is estimated for the remaining features in the present cluster. It is described as follows:

and choose another feature. When □, the ant leaves the current cluster and goes to another cluster. The parameter is used to switch among exploitation and exploration phase and the threshold value □is utilized to control the number of features to be selected within a cluster. The above process is continued until selecting features from all the available clusters. After going through all the clusters, the features selected by first ant are stored and the next ant enters into feature selection process. The same procedure is repeated in a cyclic order for the required amount of iterations. The pheromone values of the features are maintained by Multiple Then, the next suitable feature is chosen on the basis of roulette wheel rule.

Step 4:On every iteration , the pheromone amount of the th feature is updated based on MDA as follows: As explained by Huang et al. [50], the output matrix of the hidden layer of neural network with th column of being the th output of hidden neuron with respect to the input variables, Further, they showed that the hidden layer bias and input weights of SLFNS are not expected to be modified and are provided in a random

Where represents the feature selected on th iteration by the th ant, denotes the number of ants and is defined manner. With this assumption, the output weights are analytically estimated through least square solution of the as the separability indexof th selected subset in the thlinear system, :

iteration, i.e., and are the between and within scatter matrices respectively, is the transform matrix from the

-dimentionalspace to the -dimentional space, where is the number of the features selected by the th ant in the th iteration and is an integer value between 1 to with as the total number of classes.

Step 5:After completing the overall iterations(, the value of pheromone is utilized to select the needed optimal feature set. In each cluster, the features are sorted according to the amount of pheromone content and the first set of features from each cluster is chosen for further processing. Therefore, for clusters, features are selected.

B. Extreme Learning Machine based classification

A brief description about classification using Extreme Learning Machine is explained in this section. Consider a training set where and denotes the input feature vector of size and target vector of size respectively. The conventional SLFNs hold an activation function and the amount of hidden neurons can be mathematically framed as follows:

Where and represent the weight vectors among the input layer and output layer of th neuron in the hidden layer respectively. is the bias of the th neuron in the hidden layer and is the target vector of the th input data. The inner product of and be . If it is possible for the SLFNs to approximate the samples with zero errors, there will be i.e., and there exists , , such that , = , ,..., . The above Equation can be reformulated as follows:

Where, min The above Equation can be easily determined by a generalized linear approach like Moor-Penrose (MP) by finding the inverse of , as is shown inthe Equation given below.

Where is the generalized inverse matrix obtained from MP approach. Utilizing this generalized inverse may result in minimized solution for the resulting least square norms. It yields the unique as well as smallest least square norms compared to existingleast square solutions. After performing effective analysis, Huanget al. [49] explained that the generalized inverse of MP obtains better ELM performance with dramatically improved learning speed.The learning process of ELM is proceeded as follows: Initially, consider a training set, , , an activation element , and total hidden neurons . (a) Randomly allocate the input weights and bias (b) Evaluate the output of Hidden layer matrix . (c) Estimate the resultant weight .

V. RESULTS AND DISCUSSION

ISBN 978-93-88808-61-3

The implementation and performance analysis of the proposed work is performed on MATLAB R2018a software running on windows operating system with 1.7 GHz CPU and 4.00 GB of RAM. This analysis is performed to know about the functioning of feature selection as well as classification approaches. The extreme learning machine is built on the basis of 10-fold cross validation process on the thyroid disease dataset.

Experimental Design

A. Feature selection

The retrospective analysis of thyroid disease is performed on pathologically verified thyroid nodules with the help of different characteristics of US images. The thyroid dataset taken for evaluation consists of1427 nodules with 1180 benign nodules and 247 malignant nodules. The benign nodules are considerably lengthier than the malignant nodules. For evaluating the performance of feature selection algorithm, different characteristics of US images were considered. It includes different features like demographic information, boundary, echo pattern, posterior acoustic pattern, margin, orientation, position, thyroid shape, tumor and size and calcification. These features are extracted from the US images. After applying GCACO algorithm, the important features are selected from the overall available features as shown in Figure 2 and Table 1.

The definitions o selected features are provided below.

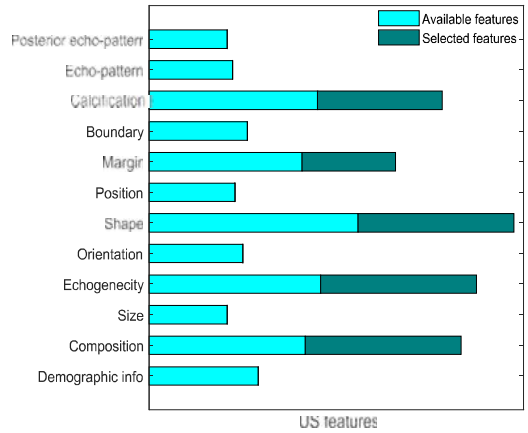


Fig 2. Features selected from the thyroid dataset by using GCACO approach

Table 1. Feature subset selected by GCACO algorithm

Size of subset	Features selected on each iteration
1	{Calcification}
2	{Calcification, Composition}
3	{Calcification, Composition, Echogenicity}
4	{Calcification, Composition, Echogenicity, Shape}
5	{Calcification, Composition, Echogenicity, Shape, Margin}

- **Calcification.**Calcification is categorized into three types like microcalcification, macrocalcification or no calcification. Microcalcification and macrocalcification are calcification with diameter less than 1 mm and larger than 1 mm respectively. If a nodule consists of both types of these calcifications, then it is remarked as microcalcification.
- **Composition.** The proportion of fluid or soft tissue in a nodule is termed as composition. It may be solid or liquid or cystic. Solid is comprised of soft tissues with liquid lesser than 10%. Predominantly solid substances are consisted of >10% liquid on<50%volume of the nodule. Incase of cystic composition, the nodule is fully or almost fully filled with liquid. One special appearance of composition is spongiform appearance that resembles like minute cystic spaces detached by thin pieces of septa.
- **Echogenicity.** In solid portions, the echogenicity is classified as iso/hyper-echogenicity, hypo-echogenicity and marked hypo-echogenicity. If the echogenicity in the nodule looks similar to the thyroid parenchyma present in their surroundings, it is termed as iso-echogenicity. If the echogenicity is low as compared to that found in strap muscles, it is called as marked hypo-echogenicity.
- **Shape.** The shape of the thyroid gland may be oval or round and it is either taller than wide or taller than long. The shape of the nodules is identified from the diameter of an anteroposterior nodule. If the anteroposterior diameter is smaller than the diameter of longitudinal and transverse planes, then the shape is said to be oval shape. Otherwise, if the anteroposterior diameter is equal to the diameter of longitudinal and transverse planes, then it is called as round shape. If the ratios of anteroposterior to transverse and longitudinal diameters are greater than one, then the structure of nodule is taller than wide and taller than long respectively.
- **Margin.** The outline of thyroid nodule is called as margin of the nodule. The margin of the nodule takes different structures like smooth margin, ill-defined margin, irregular margin and microlobulated.

Table 2. Ultrasound features of thyroid nodules

Features	Number of benign nodules (n=1180)	Number of malignant nodules (n=247)	*p*-value (analysis)
Calcification			<0.001
Macro-calcification (n=303)	257	46	
Micro-calcification (n=127)	50	77	
No calcification (n=997)	873	124	
Composition			<0.001
Solid (n=983)	763	220	
Mixed (n=444)	417	27	

Echogenicity			<0.001
Hyper-echogenicity (n=854)	650	204	
Hypo-echogenicity (n=412)	370	42	
Marked hypo-echogenicity (n=161)	160	1	
Margins			<0.001
Smooth (n=1010)	985	25	
Microlobulated (n=243)	178	65	
Irregular (n=174)	17	157	
Shape			<0.001
Wider than tall (n=1253)	1117	136	
Taller than wide (n=174)	63	111	

The information about the extracted features is given in Table 2.These features are significant to identify the risk of malignancy associated with thyroid nodules. From Chi- square analysis, the nodules with the following criteria are considered as malignant

Extreme Learning Machine for Thyroid Nodule Classification with Graph Cluster Ant Colony Optimization Based Feature Selection

Table 3. General structure of the confusion matrix

Type of classifier		**Predicted**	
		Malignant nodules	**Benign nodules**
Actual	Malignant nodules	True positive	False Negative
	Benign nodules	False Positive	True Negative

The equations for calculating accuracy, sensitivity and specificity from the confusion matrix are described above. The region under the receiver operating characteristic curve is termed as the area under curve which draws the true positives rather than the false positive rates. Finally, the classifier with higher AUC is remarked as the best classifier than the classifier that produces smaller AUCs. A classifier with AUC equal to one is concluded as a perfect classifier.

B. Performance Analysis

The performance of the proposed thyroid disease nodule classification method is evaluated by analyzing the functions of both GCACO feature selection and extreme machine learning classification algorithms. The effectiveness of classification based on the selected features is compared with existing methods such as L-Score, F-Score, ReliefF and UFSACO. Table 4-6 shows the comparison table for classification accuracy, sensitivity and specificity obtained with different feature selection algorithms. Different feature criteria that decides the efficiency of classification includes composition, calcification, margin, shape, solid part of echogenicity and size of nodules that are equal to or larger than 5 mm.

The comparison of classification accuracy of GCACO with existing feature extraction methods such as L-score, F-score, ReliefF and UFSACO are shown in Table 4. From this analysis, it is clear that GCACO produce better accuracy in classification than the existing methods. The maximum classification accuracy obtained for GCACO is 98%, 95% and 97% with feature subsets 1, 4 and 5 respectively. The featuresubsets extracted from relief methods also produced better results with 96% and 95% accuracy for number of features 2 and 3 respectively. On the other hand, L-score, F- score and UFSACO produced worst results on classification accuracy. The classification sensitivity for GCACO outperformed the existing methods with 98%, 96% and 94% sensitivity with subsets 1, 3 and 4 respectively. It is listed in Table 5. The ReliefF method utilized 2 and 5 number of featuresfor performing the classification. The sensitivity of these featuresubsets are found to be 97% and 95% respectively. As shown in Table 6, the classification specificity of GCACO for the subsets 1, 2, 3 and 4 is found to be 99%, 95%, 98% and 95%

Table5. Comparison of classification sensitivity with different feature selection algorithms

No. of features in feature subset	**L-score**	**F-score**	**ReliefF**	**UFSACO**	**Proposed**
1	86.11±6.56	89.35±1.52	94.67±2.77	95.35±3.45	**98.56±1.24**
2	89.46±4.38	95.67±4.45	**97.46±1.56**	86.56±2.35	93.67±4.16
3	84.46±2.45	90.78±7.54	90.67±3.86	88.67±4.26	**96.67±4.24**
4	94.45±6.35	94.24±7.87	91.35±7.45	92.57±4.26	**94.46±1.23**
5	95.47±3.88	95.24±1.48	**95.67±2.56**	94.67±3.54	94.57±2.65

respectively. But, L-score produced 91% better classification specificity with 4 features in the feature subset. Thus, compared to univariate feature selection approaches like L-score, F-score and Relief F, the proposed multivariate feature selectionapproach functions significantly better. The performance of proposed ELM classifier is evaluated by comparing it with other classification techniques such as SVM and KNN. The average accuracy, sensitivity and specificity obtained for all these techniques for the features of

thyroid datasets are presented in Table 5. From there, it is observed that the ELM classifier outperforms the existing approaches used for classifying the thyroid dataset with selected feature subsets. Further, the improved accuracy, sensitivity and specificity resemble the higher grade of relevancy and redundancy between the selected features from the input dataset.Thus the significant performance obtained from the ELM classifier is due to high relevancy as well as low redundancy among the features considered for classification. The confusion matrix outcomes of three classifiers such as ELM, ANN and SVM are presented in Table 7. It is identified that ELM has properly differentiated 57 malignant nodules and 87 benign nodules. Further, it misidentified 16 malignant nodules as benign and 10 benign nodules as malignant. On the other hand, ANN perfectly classified 49 malignant nodules and 23 benign nodules. In addition, it misjudges 23 malignant nodules as benign and 17 benign nodules as malignant. Finally, the SVM classifier predicted 51 malignant nodules and 83 benign nodules. However, it misjudges 20 malignant nodules as benign and 14 benign nodules as malignant. This visualize that ELM outperforms the other two methods in differentiating the type of nodules.

Table 7. Confusion matrix of ELM

	Type of classifier	Predicted nodules	
		Malignant	Benign
Actual nodules	**ELM classifier**		
	Malignant	57	16
	Benign	10	87
	ANN classifier		
	Malignant	49	23
	Benign	17	80
	SVM classifier		
	Malignant	51	20
	Benign	14	83

The performance of ELM based on different hidden neurons is shown in Figure 3. As shown, the accuracy and specificity fluctuates through certain limit and remains stable as the total amount of neurons is increased. The sensitivity and AUC is found to be almost stable throughout the hidden neurons. At 20 neuron case, the accuracy of the extreme learning machine classifier is found to be higher and so, it is taken as the optimal number of the neurons.

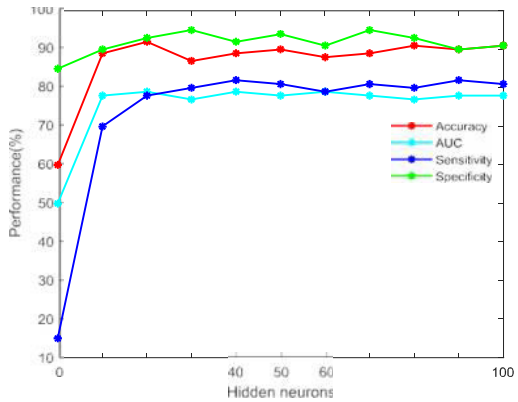


Figure 3. Performance of ELM based on the hidden neurons

The comparative results for accuracy, AUC, sensitivity and specificity of ELM and existing techniques like SVM and ANN classifiers are shown in Figure 4.It is provided in terms of the mean value of different parameters taken for analysis. The figure depicts that the performance of ELM is better than other machine learning classification techniques such as SVM and ANN. Further, the performance of SVM is slightly higher than that of the artificial neural network classifier. This describes that that ELM is the best method to be used for analyzing the thyroid disease using US characteristics.

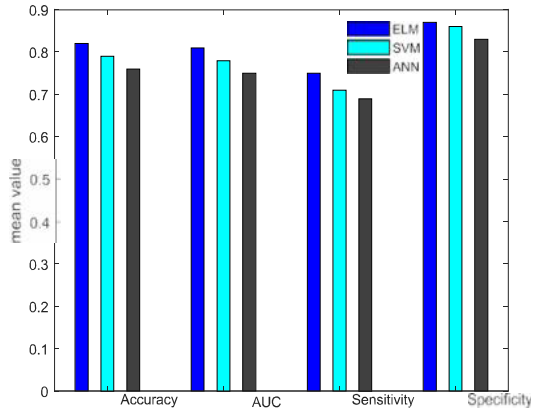


Figure 4. Comparative analysis of classifier algorithms in terms of: accuracy, AUC, sensitivity and specificity

VI. CONCLUSION

The ultrasound is defined as a non-invasive tool that is utilized for the diagnosis of thyroid lesions because of its affordable price and ease of availability. During thyroid treatment, differentiation of malignancy nodules remain a critical task due to the challenges faced by texture analysis and employing machine learning algorithms in modern diagnostic procedures. Thus, it is necessary to modernize and authorize these processes for the purpose of making it to be desire in the field of thyroid diagnostics. In this paper, Graph- Clustering Ant Colony Optimization based Extreme Machine Learning approach is introduced for the detection of malignancy risk associated with thyroid nodules. Using GCACO, the features are partitioned into group of clusters and the clusters are assembled as an undirected graph with community detection algorithm. Afterwards, ACO is employed to select the optimal features from the group of clusters. In this work, five of the ultra-sonographic features are taken as discriminant features from the US thyroid dataset. The GCACO feature selection method is multivariate and it is compared with existing univariate methodologies like L-score, F-score, ReliefF and UFSACO. It is identified that GCACO significantly identifies the suitable features and functions better than that of the univariate algorithms introduced for the same purpose. Further, the simulation results demonstrate that ELM classifier accurately differentiates the malignant

nodules from benign nodules. Thus, GCACO based ELM classifier can be efficiently applied for clinical diagnosis of thyroid disorders and produce effective result in thyroid treatment.

REFERENCES

1. Acharya, U.R., Chowriappa, P., Fujita, H., Bhat, S., Dua, S., Koh, J.E., Eugene, L.W.J., Kongmebhol, P. and Ng, K.H., 2016. Thyroid lesion classification in 242 patient population using Gabor transform features from high resolution ultrasound images. Knowledge-Based Systems, 107, pp.235-245.
2. Ardakani, A.A., Gharbali, A. and Mohammadi, A., 2015. Application of texture analysis method for classification of benign and malignant thyroid nodules in ultrasound images. Iranian journal of cancer prevention, 8(2), p.116.
3. Bakshi, N.A., Mansoor, I. and Jones, B.A., 2003. Analysis of inconclusive fine-needle aspiration of thyroid follicular lesions. Endocrine pathology, 14(2), pp.167-175.
4. Chang, C.Y., Chen, S.J. and Tsai, M.F., 2010. Application of support- vector-machine-based method for feature selection and classification of thyroid nodules in ultrasound images. Pattern recognition, 43(10), pp.3494-3506.
5. Chen, F.L. and Ou, T.Y., 2011. Sales forecasting system based on Gray extreme learning machine with Taguchi method in retail industry. Expert Systems with Applications, 38(3), pp.1336-1345.
6. Choi, W.J., Park, J.S., Kim, K.G., Kim, S.Y., Koo, H.R. and Lee, Y.J., 2015. Computerized analysis of calcification of thyroid nodules as visualized by ultrasonography. European journal of radiology, 84(10), pp.1949-1953.
7. Cooper, D.S., Doherty, G.M., Haugen, B.R., Kloos, R.T., Lee, S.L., Mandel, S.J., Mazzaferri, E.L., McIver, B., Sherman, S.I. and Tuttle, R.M., 2006. Management guidelines for patients with thyroid nodules and differentiated thyroid cancer: The American Thyroid Association Guidelines Taskforce. Thyroid, 16(2), pp.109-142.
8. Erdem, G., Erdem, T., Muammer, H., Mutlu, D.Y., Fırat, A.K., Sahin,
   I. and Alkan, A., 2010. Diffusion- weighted images differentiate benign from malignant thyroid nodules. Journal of Magnetic Resonance Imaging, 31(1), pp.94-100.
9. Frates, M.C., Benson, C.B., Doubilet, P.M., Kunreuther, E., Contreras, M., Cibas, E.S., Orcutt, J., Moore Jr, F.D., Larsen, P.R., Marqusee, E. and Alexander, E.K., 2006. Prevalence and distribution of carcinoma in patients with solitary and multiple thyroid nodules on sonography. The Journal of Clinical Endocrinology & Metabolism, 91(9), pp.3411-3417.
10. Gomathi, M. and Thangaraj, P., 2010. A computer aided diagnosis system for lung cancer detection using support vector machine. American Journal of Applied Sciences, 7(12), p.1532.
11. Han, F., Huang, D.S., Zhu, Z.H. and Rong, T.H., 2006. The forecast of the postoperative survival time of patients suffered from non-small cell lung cancer based on PCA and extreme learning machine. International journal of neural systems, 16(01), pp.39-46.
12. Helmy, T. and Rasheed, Z., 2009, May. Multi-category bioinformatics dataset classification using extreme learning machine. In 2009 IEEE Congress on Evolutionary Computation (pp. 3234-3240). IEEE.
13. Huang, G.B., Zhu, Q.Y. and Siew, C.K., 2004. Extreme learning machine: a new learning scheme of feedforward neural networks. Neural networks, 2, pp.985-990.
14. Kabir, M.M., Shahjahan, M. and Murase, K., 2011. A new local search based hybrid genetic algorithm for feature selection. Neurocomputing, 74(17), pp.2914-2928.
15. Li, L.N., Ouyang, J.H., Chen, H.L. and Liu, D.Y., 2012. A computer aided diagnosis system for thyroid disease using extreme learning machine. Journal of medical systems, 36(5), pp.3327-3337.
16. Li, X., Zhang, S., Zhang, Q., Wei, X., Pan, Y., Zhao, J., Xin, X., Qin, C., Wang, X., Li, J. and Yang, F., 2019. Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. The Lancet Oncology, 20(2), pp.193-201.
17. Li, X., Zhang, S., Zhang, Q., Wei, X., Gao, M., Zhang, W. and Chen, K., 2019. Deep convolutional neural network models for the diagnosis of thyroid cancer–Authors' reply. The Lancet Oncology, 20(3), p.e131.
18. Liu, N., Lin, Z., Koh, Z., Huang, G.B., Ser, W. and Ong, M.E.H., 2011. Patient outcome prediction with heart rate variability and vital signs. Journal of Signal Processing Systems, 64(2), pp.265-278.
19. Koundal, D., Gupta, S. and Singh, S., 2018. Computer aided thyroid nodule detection system using medical ultrasound images. Biomedical Signal Processing and Control, 40, pp.117-130.
20. Ma, J., Wu, F., Zhu, J., Xu, D. and Kong, D., 2017. A pre-trained convolutional neural network based method for thyroid nodule diagnosis. Ultrasonics, 73, pp.221-230.
21. Moradi, P. and Rostami, M., 2015. Integration of graph clustering with ant colony optimization for feature selection. Knowledge-Based Systems, 84, pp.144-161.
22. Ouyang, F.S., Guo, B.L., Ouyang, L.Z., Liu, Z.W., Lin, S.J., Meng, W., Huang, X.Y., Chen, H.X., Qiu-gen, H. and Yang, S.M., 2019. Comparison between linear and nonlinear machine-learning algorithms for the classification of thyroid nodules. European Journal of Radiology, 113, pp.251-257.
23. Shankar, K., Lakshmanaprabu, S.K., Gupta, D., Maseleno, A. and de Albuquerque, V.H.C., 2018. Optimal feature-based multi-kernel SVM approach for thyroid disease classification. The Journal of Supercomputing, pp.1-16.
24. Sivagaminathan, R.K. and Ramakrishnan, S., 2007. A hybrid approach for feature subset selection using neural networks and ant colony optimization. Expert systems with applications, 33(1), pp.49- 60.
25. Sollini, M., Cozzi, L., Chiti, A. and Kirienko, M., 2018. Texture analysis and machine learning to characterize suspected thyroid nodules and differentiated thyroid cancer: Where do we stand?. European journal of radiology, 99, pp.1-8.
26. Tabakhi, S., Moradi, P. and Akhlaghian, F., 2014. An unsupervised feature selection algorithm based on ant colony optimization. Engineering Applications of Artificial Intelligence, 32, pp.112-123.
27. Tsantis, S., Dimitropoulos, N., Cavouras, D. and Nikiforidis, G., 2009. Morphological and wavelet features towards sonographic thyroid nodules evaluation. Computerized Medical Imaging and Graphics, 33(2), pp.91-99.
28. Wu, Y., Yue, X., Shen, W., Du, Y., Yuan, Y., Tao, X. and Tang, C.Y., 2013. Diagnostic value of diffusion-weighted MR imaging in thyroid disease: application in differentiating benign from malignant disease. BMC medical imaging, 13(1), p.23.
29. Yong, Z., Dun-wei, G. and Wan-qiu, Z., 2016. Feature selection of unreliable data using an improved multi-objective PSO algorithm. Neurocomputing, 171, pp.1281-1290.
30. Zhang, R., Huang, G.B., Sundararajan, N. and Saratchandran, P., 2007. Multicategory classification using an extreme learning machine for microarray gene expression cancer diagnosis. IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), 4(3), pp.485-495.

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

Advanced Machine Learning Techniques To Handle Brain Image Segmentation And Tumor Classification Over Bio-Medical Images

Mr .M AHARONU
Assistant Professor
Dept of CSE
Malla Reddy College Of Engineering
Hyderabad
Email:aharon.mattakoyya@gmail.com

Mrs. V Devi Priya
Assistant Professor
Dept of CSE
Malla Reddy College Of Engineering
Hyderabad
Email:vaddi.devipriya@gmail.com

Abstract

In real time applications, to evaluate mathematical closely related relations Rough set theory, Fuzzy set theory and rough set theory are the mathematical linear tools for uncertain data elements. Some of the researchers introduced rough sets, rough sets and fuzzy set by connected and combining all set theories together. In this research, we discuss about different combined notations of fuzzy, rough and rough set theories, and also discuss basic methods used to describe about above set theories effectively. We present the concepts related to rough based Rough intuitionistic fuzzy sets, intuitionistic fuzzy rough sets and discuss about basic properties of those set theories effectively. Furthermore, we discuss about classical presentation of rough based intuitionistic fuzzy sets in detail with approximate operations in real time synthetic applications. Segmentation of magnetic resonance images is medically complex and important for study and diagnosis of medical brain images, because of its sensitivity in terms of noise for brain medical images. These are the main issues in classification of brain images. Because of uncertainty & vagueness of brain medical images, so that rough sets, fuzzy sets and Rough sets are mathematical tools evaluate and handle uncertainty and vagueness in medical brain images. Traditionally, different type of fuzzy sets, Rough sets and rough sets based approaches were introduced, they have different several drawbacks with respect to different parameters. This research introduces a novel image segmentation (Classification) calculation method i.e. Enhanced and Explored Intuitionistic Rough based Fuzzy C-means Approach (EEISFCMA) with Support Vector machine classifier to estimation of weight bias parameter for brain image segmentation. Intuitionistic

Rough b ased fuzzy sets are generalized form of fuzzy, rough sets and their representative elements are evaluated with non-membership and membership value. Proposed algorithm of this paper consists standard features of existing clustering without spatial weight context data, it defines sensitive of noise in brain images, so that our proposed algorithm deals with intensity and noise reduction of brain image effectively. Furthermore, to reduce iterations in clustering, proposed algorithm initializes cluster centroid based on weight measure using max-dist evaluation method before execution of proposed algorithm. Experimental results of proposed approach carried out efficient image segmentation results compared to existing segmented approaches developed in brain image and other related images. Mainly proposed approach have consists better experimental evaluation based on results.

1. Introduction

In recent times, the introduction of information technology and e-health care system in the medical field helps clinical experts to provide better health care to the patient. Brain tumors affect the humans badly, because of the abnormal growth of cells within the brain. It can disrupt proper brain function and be life-threatening. Two types of brain tumors have been identified as benigntumors and malignanttumors. Benign tumors are less harmful than malignant tumors as malignant are fast developing and harmful while benign are slow growing and less harmful. The various types of medical imaging technologies based on noninvasive approach like; MRI, CT scan, Ultrasound, SPECT, PET and X-ray [1]. When compared to other medical imaging techniques, Magnetic Resonance Imaging (MRI) is majorly used and it provides greater contrast images of the brain and cancerous tissues. Therefore, brain tumor identification can be done through MRI images [2]. This paper focuses on the identification ofbrain tumor using image processing techniques. The detection of a brain tumor at an early stage is a key issue for providing improved treatment. Once a brain tumor is clinically suspected, radiological evaluation is required to determine its location, its size, and impact on the surrounding areas. On the basis of this information the best therapy, surgery, radiation, or chemotherapy, is decided. It is evident that the chances of survival of a tumor-infected patient can be increased significantly if the tumor is detected accurately in its early stage [3]. As a result,

the study of brain tumors using imaging modalities has gained importance in the radiology department. In this paper the brain tumor identification is done by an image processing. In this paper, there are four process are done to identify the brain tumors. The first process is pre processing the image data from the collection of database using median filtering, second stage is segmentation using Fuzzy C-means Clustering Algorithm [4], third stage is feature extraction using Gray Level Co-Occurrence Matrix (GLCM), [5] and the fourth stage is classification using ensemble classifiers is the combination of neural network, Extreme Learning Machine (ELM) and Support Vector Machine classifier (SVM). This will be discussed briefly in this following section.

2. Literature Survey

We can observe different data sets like Fuzzy sets, Rough sets and Soft sets notations with mathematical evaluations in real time applications based on different theories and developments. Present day's brain image segmentation is the basic problem to evaluate brain tumor decease in artificial intelligence real time applications. In medical image processing applications, brain tumor detection is a challenging task for real time medical applications. Traditionally some of the research authors introduced different machine learning methods, clustering approaches, classification approaches and filtering approaches to evaluate the basic procedure of the brain image segmentation in both theoretical and practical implementations based on above discussed data sets. All those approaches have some cons and pons in their implementations. In this section, we discuss about each technique implementations using real time data sets in image segmentation. Table 1 gives the brief discussion about all those techniques

Segmentation Approach	**Author**	**Description**	**Advantages**	**Disadvantages**
Adaptive Threshold	S. Jansi et.al	Based on image background, divide image into different dynamic regions based on threshold of various pixel	It will be worked based on thresholds	Less accuracy when rotation of different images applied, High time for processing images

		values		
K-Means Clustering	D.Selvaraj et.al	K-means clustering algorithm worked based on geometric interpretation of data. Based on centroid in images, it can identify brain tumor in images.	Less time for processing brain tumor segmentation, It is iterative process.	Less Accuracy, and Less false positive rate, not worked for large scale datasets
Improved K-Means Clustering	P. Vijayalakshmi et.al	Based on initial presentation of clustering identify brain tumor pixels in image segmentation.	It is easiest process, More accurate and high resolution	Less sensitivity and high time for image segmentation.
Fuzzy C-Means Clustering	M. Rakesh et.al	Based on given and pre-defined region and based on similarity measure identify brain tumor identification in images	More accurate in image segmentation	Give more time to identify tumor in brain images
Adaptive Fuzzy K-means Clustering	S. N. Sulaiman et.al	Based on degree measure relationship in images to identify brain tumor.	It is used to process Magnetic Resonance Images (MRI) Images	It is not applicable for qualitative and quantitative MRI brain images.
Region growing	Sudipta Roy et.al	Brain tumor identification is processed based on kindly segmentation	Extraction surface points may cardiac segmentation of	Requires user interface to formulate selection tumor presented surface from

ISBN 978-93-88808-61-3

		process applied on medication images	images	segmented images
Mean shift	Vishal B et.al	It is computer vision based non parametric clustering approach in medical image processing	It detect brain tumor on n-dimensional set presentation	Because of iterations in real time presentations, it computes high time complexity in segmentation
Watershed segmentation	Deorah et al	To identify foreground and background in image segmentation	Capturing of weak pixel formation in image segmentation, Less time for segmentation	Selection of Seed point selection is low, Increase convergence rate.
Level Set Model	Jiang Zhang et.al	To identify brain tumor in images based on surfaces at each dimension	Detection occurred based on level of surface identifications	It is not worked properly if curve was breaking.
K-Nearest Neighbour	Warfield et.al	Instance based brain tumor detection in brain image segmentation procedures	It is simplest approach to identify image segmentation, Increase accuracy	Statistical model to identify brain tumor presentations in brain images.
Support Vector Machine	Vapnik et.al	It is a supervised machine learning procedure to identify brain tumor presentation in image segmentation	It is an attractive and symmetric method to detect brain tumor image segmentation	Accuracy is very low in classification

Principal Component Analysis	Sumitra et al.	Based on principle feature presentation in images, identify the brain tumor classification in image segmentation	Reduce the large dimensionality in image segmentation	Less decomposition rate in image segmentations
Expectation maximization	Moon et al.	Based on some previously available tumor rules identify detection in brain image segmentation	Differentiate healthy and timorous tissues in brain image segmentation	It have intensity distribution of brain images.
Hierarchical clustering	Kshitij et al	Based on grouped tree clustering, identify tumors in brain images.	Accuracy is very high	Time complexity is very low
Back Propagation Algorithm	Rumelhard, D et.al	This method works properly in feed forward network ro identify tumor in brain images	Time complexity is less and easily verifiable	Less accuracy with feature extraction based on signal waves

Motivation

Consider the preliminaries present in table 1, we focus on development of advanced techniques to identify brain tumor in brain images based on segmentation/other properties. Our research mainly implement false positive rate, less time complexity and increase the accuracy in brain image segmentation to get better performance results of detection brain tumor in brain images.

3. Problem Statement

Fuzzy sets, soft sets and rough sets are the effective data processing frameworks for decision making relative to information

processing systems, information retrieval and other conclusive relations present in data, especially in some types of uncertain data events. So it is an efficient concept to process and effective dealing to evaluate uncertain data with different parameters. Consistently, number of researchers or authors has been introduced number of techniques in practical and theory oriented applications.

Define and discuss about different concepts related to fuzzy sets, rough sets and soft sets theories with their implementation in various fields with existing literature. To further implementation of this work is to develop soft rough with intuitionistic fuzzy sets to generalize properties of real time applications like image segmentation in brain oriented applications. We extend our research to support different mathematical evaluations in uncertain data processing in brain image segmentation with practical implementation.

4. Proposed Methodology

Main objective parameters for defined function i.e. K(X, Y,) to minimize standard representations for c-means for image segmentation in brain medical images. First we take derived parameters of defined function K(X, Y,) with respect to membership parameters x_{ij}, cluster centroid v_i and biased field $_k$ setting them into 0 and results of estimation matrixes of X(membership matrix), Y(centroid matrix) and (bias matrix). Based on these estimated results, we form our novel calculation and compute the classification of tissue and bias function field. Newly generated function of proposed approach is

$$K(X, Y,) = \sum_{i=1}^{c}\sum_{k=1}^{n} x_{ik}^{m} \| g_k x_k - y_i \|^2 x_{ik}^{m}$$

Where $x_k = \frac{g_k}{n}$

Estimation of Bias field

Taking the derivative of K(X, Y,) with respect to $_k$ and assign them into 0 then we have

$$\frac{K(X, Y,)}{_k} = \sum_{i=1}^{n} x_{ik}^{m}(g_k x_k - y_i)^2$$

Second summation of k^{th} term with respect to $_k$ then us have the following expression

$$\sum_{i=1}^{c} x_{ik}^{m} g_k \sum_{i=1}^{c} x_{ik}^{m} x_k - \sum_{i=1}^{c} x_{ik}^{m} v_{ik} = 0$$

Differentiating the distance expression, then we obtain following expression

$$g_k \sum_{i=1}^{c} x_{ik}^{m} x_k - \sum_{i=1}^{c} x_{ik}^{m} v_{ik} = 0$$

ISBN 978-93-88808-61-3

$$g_k \; x_{ik}^m \; g_k \; x_{ik}^m \; x^m v$$

Distance based gradient function for bias field function is as follows:

$$\frac{1}{k} \quad y_k \quad \frac{\sum_{i=1}^{c} x_{ik}^m v_k}{\sum_{i=1}^{c} x_{ik}^m}$$

If $(0,1)$ is the weight of membership function, then generated bias data is 0.007 and increase this from 0.001 to $_2, _3, \ldots, _{10}$.

Updated Centroid of Cluster

Again taking the derivative of K(X, Y,)

with respect y_i and setting results is zero, then generated function is

$$\sum_{k=1}^{n} x_{ik}^m (g_k \; _k \; _k \; y_i) \; 0$$

Where $y_i^* = \frac{\sum_{k=1}^{n} x_{ik}^m (g_k \; _k \; _k)}{\sum x_{ik}^m}$ after solving the above equation.

Intuitionistic Fuzzy based Image Representation

Intuitionistic fuzzy sets [IFS] representation of image for image segmentation. The presented image consists N*M size and the

, where i is in between1to N*M, then image X to be represented in IFS as follows:

$$X \; \{ a_i, (a_i), (a_i)) \mid a_i \; A\}$$

with $(a_i) \; 1 \; ((a_i) \; (a_i))$, (a_i) where

is membership function and (a_i) is non-member function and (a_i) is the mean

pixel value of image. After evaluating fuzzy image representation update each cluster based on different pixel values of image.

Evaluation of Membership

After minimize the above equation, with different constraints using Lagrange multiplier calculation

$$L_u(X, Y,) \; \sum^{c} \sum^{n} x^m \| g_k \; _k \; _k \; y_i \|^2 \; x \; 1 \; \sum^{c} x_{ik}^m \; 1 \; \sum^{c}$$

After taking derivative of $L(X, Y,)$ with respect to x_{ik} and set result into zero, then we have

$$\frac{K(X, Y,)}{} \; mx_{ik}^{m-1} \| g_k \; _k \; _k \; y_i \|^2 \; mx_{ik}^{m-1} \; x_{ik} \; x_{ik}^* \; 0$$

After solving the above equations based on different parameters for membership

parameter sequences can be re=written as follows:

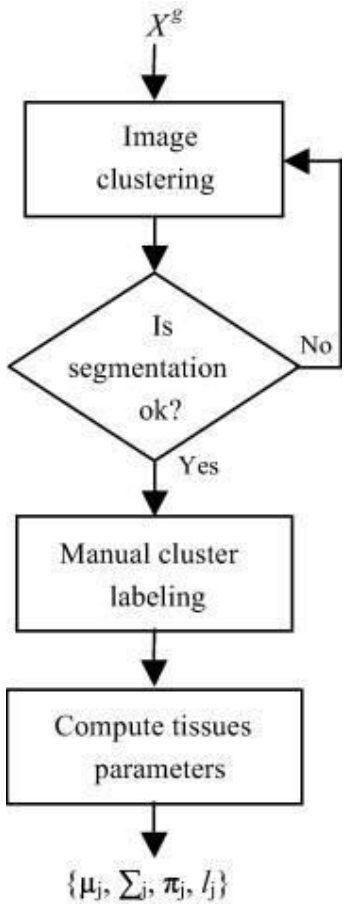
$$x_{ik}^* = \sum_{j=1}^{c} \left(\frac{\| g_k \; _k \; _k \; y_i \|^2 \; x^{-\frac{1}{m-1}}}{\| g_k \; _k \; _k \; y_j \|^2 \; x} \right)^{-1}$$

In the above equation c is number of

value of each pixel in image i.e.$A \square \{a_i / a_i\}$ centroid; g is gain function andis constant for membership function with different parameters.

EEISFCMA is evaluated on publicly available brain images, for example we collected brain images from https://www.nitrc.org/frs/?group_id=48&release_id=3124 and http://brainweb.bic.mni.mcgill.ca/brainweb/ with simulated brain image databases. We download these images from web urls and then convert into Matlab readable format and then we can pre-process for feature extraction to segment images using readable Rough ware i.e analysis and visualization of image. Proposed approach can be implemented in Latest Mat lab version with latest system configurations and this section describes implemented results. This section describes experimental results of different traditional approaches like k-means, fuzzy c-Mean, Generalized Fuzzy C-means, Gaussian Kernel based Fuzzy c-Means algorithm (GKFCM) and Rough fuzzy rough sets c-means (SFRCM) with proposed approach at segmentation accuracy and jacquard co-efficient for brain segmented images.

5. System Design



Design implementation of brain image segmentation for bio-medical images from different sources.

6. References

[1] Anupama Namburu, Srinivas kumar Samay, Srinivasa Reddy Edara, "Rough fuzzy rough set-based MR brain image segmentation",Proceedings in Applied Rough Computing xxx (2016) xxx–xxx. © 2016 Elsevier B.V. All rights reserved.

[2] S. Ramathilagam, R. Pandiyarajan, A. Sathya, R. Devi, S.R. Kannan," Modified fuzzy c-means algorithm for segmentation of T1–T2-weighted brain MRI", © 2010

Elsevier B.V. All rights reserved , Journal of Computational and Applied Mathematics 235 (2011) 1578–1586.

[3] K.V. Leemput, F. Maes, D. Vandermeulen, P. Suetens, Automated model based bias field correction of MR images of the brain, IEEE Trans. Med. Imaging 18 (1999) 885–896.

[4] M.N. Ahmed, N.A. Mohamed, A.A. Farag, T. Moriarty, A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data, IEEE Trans. Med. Imaging 21 (2002) 193–199.

[5] R.L. Cannon, J.V. Dave, J.C. Bezdek, Efficient implementation of the fuzzy c-means clustering algorithms, IEEE Trans. Pattern Anal. Mach. Intell. PAMI-8 (2) (1986) 248–255.

[6] Zujun Hou, A review on MR image intensity in homogeneity correction, Int. J. Biomed. Imaging (2006) 1–11. Article ID 49515.

[7] B.R. Condon, J. Patterson, D. Wyper, Image nonuniformity in magnetic resonance imaging: its magnitude and methods for its correction, Br. J. Radiol. 60 (1987) 83–87.

[8] M. Tincher, C.R. Meyer, R. Gupta, D.M. Williams, Polynomial modelling and reduction of spatial body-coil spatial in homogeneity, IEEE Trans. Med. Imaging 12 (1993) 361–365.

[9] S. Lai, M. Fang, A new variational shape-from orientation approach to correcting intensity inhomogeneities in MR images, in: Proc. of Workshop on Biomedical Image Analysis CVPR98, 1998, pp. 56–63.

[10] S.E. Moyher, D.B. Vigneron, S.J. Nelson, Surface coil MR imaging of the human brain with an analytic reception profile correction, J. Magn. Reson. Imaging 5 (1995) 139–144.

[11] S. Krinidis, V. Chatzis, A robust fuzzy localinformation c-means clusteringalgorithm, IEEE Trans. Image Process. 19 (5) (2010) 1328–1337.

[12] C. Li, J.C. Gore, C. Davatzikos, Multiplicative intrinsic component optimization(MICO) for MRI bias field estimation and tissue segmentation, Magn. Reson.Imaging 32 (7) (2014) 913–923.

[13] Z. Pawlak, Rough set approach to knowledge-based decision support, Eur. J.Oper. Res. 99 (1) (1997) 48–57.

[14] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, Int. J. Gen. Syst. 17(2–3) (1990) 191–209.[15] P. Lingras, C. West, Interval set clustering of web users with rough k-means, J.Intell. Inf. Syst. 23 (1) (2004) 5–16.

[16] P. Maji, S.K. Pal, Rough set based generalized fuzzy c-means algorithm andquantitative indices, IEEE Trans. Syst. Man Cybern. B 37 (6) (2007)1529–1540.

[17] P. Maji, S.K. Pal, RFCM: a hybrid clustering algorithm using rough and fuzzysets, Fundam. Inform. 80 (4) (2007) 475–496

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

An Iot Based Approach For Energy Flexible Control Of Production Systems

P.Sandeep[1] Assistant Professor Malla Reddy College of Engineering
B.Shiva Karthik[2] Assistant Professor Malla Reddy College of Engineering

Abstract

Due to the increasing amount of renewable energy on the energy market resulting in a higher volatility of energy supply, manufacturing companies have an enhanced awareness of their energy demand in order to benefit from alternating prices. Energy flexibility is an opportunity to adapt manufacturing systems to the changing circumstances. The idea of energy flexibility follows the approach of synchronizing energy demand with supply, e.g. to exploit alternating weather conditions. This paper presents an energy-aware demand side management (DSM) approach to control manufacturing systems on the component level. The developed closed loop control is based on an algorithm fed with manufacturing, energy and environmental data and is realized at an Internet of Things (IoT) platform. Based on machine tool models the energy demand of a hypothetical factory is simulated. Taking on-site power generation data into account, the aim of the developed energy-aware control loop is to reduce the appearing residual power that must be balanced with grid-supplied power.

Keywords: energy flexibility; machine tools; on-site power generation; Internet of Things

1. Introduction

To achieve global climate agreements recently updated at the UN conference in 2015, new restrictions addressing the greenhouse gas (GHG) emissions were introduced by the German government. The Renewable Energy Law defines feed-in remuneration to increase the amount of renewable energy. As a result, the share of renewable energy has been increasing continuously to a rate of 29 % (188 GWh) in 2016 [1].

The German climate protection plan 2050 [2] includes a holistic energy concept addressing the energy sector, buildings, transport, agriculture and industry. For the industry sector, a reduction of GHG emissions of 49 % is striven for. Both the changing energy market with an increasing share of renewables and the rising viability of on-site power generation for manufacturing companies lead to a volatile energy supply. The adaption of the energy demand to supply plays a significant role to ensure competitiveness due to process stability, product quality and cost advantages.

This paper introduces an energy-conscious demand side management approach to control manufacturing systems on the component level. Based on a machine model the performance of the developed closed loop control is analyzed concentrating on the impact on CO_2 emissions, costs and time of grid neutrality. Furthermore, factory conditions for the application of the most suitable methods are identified and the approach is realized on an IoT platform.

2. Energy flexibility in smart factories

The future factory

Due to environmental circumstances, the entire factory structure will change resulting in new challenges. The conventional goal triangle in manufacturing companies is evolveing to a pyramid with the additional targets flexibility and sustainability [3] (figure 1).

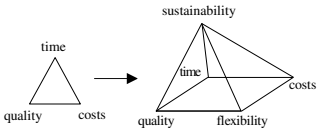


Fig. 1. Evolution of targets in manufacturing companies

Addressing sustainability, future factories should take social, economic and ecological aspects into account. A future concept of a sustainable manufacturing site is for example introduced by Stoldt et al. [4] with the key issues resource efficiency, zero emission and embedding people.

Besides the increasing awareness for sustainability, the digitalization influences the future factory significantly. The

advantages of the implementation of IoT technology in the future are to be found in literature:

- Flexibility, compatibility, scalability, ubiquity [5-8]
- Resource, cost and operational efficiency [7,8]
- Real-time capability and robustness [6,7,9]
- Usability and transparency [6,9]
- Complexity and intelligence [6,9]

The above-mentioned advantages of interlinking based on innovative information technology accelerates the fourth industrial revolution. Therefore, it is assumed, that the conventional automation pyramid will evolve to CPS (cyber-physical system) - based automation [10].

The implementation of IoT technologies supports the adaptation of future factories to changing environmental circumstances and can be useful for energy management in manufacturing companies.

Evaluation of energy flexibility in the research field of energy management in production systems

The research field of energy management includes different approaches and levels to face the challenges along with resource scarcity. Both energy data acquisition and analysis as well as energy data monitoring are requirements for energy flexible production planning or control. Overall energy management includes all aspects regarding resource allocation and planning. The evaluation of energy flexibility is observed on all re-search field levels.

In general, Reinhart et al. [11] define energy flexibility as *the capability of a production system to adapt quickly and with low financial expenditure to changes on the energy market*. Based on this definition, dimensions to identify the energy flexibility on the machine level are introduced [12]. Accordingly, energy flexible machines have low switching times, high power change rates and short critical times. Popp et al. [13] determine the degree of technical energy flexibility based on the components¶ demand and their relation among each other quantified with the Energy Independency Indicator (EII). Furthermore, energy flexibility indices are defined to evaluate energy flexibility on the component and on the machine level [14]. Simon et al. [15] introduce a method for the technical and economical evaluation of energy flexibility regarding the identification and categorization of measures.

The introduced evaluation approaches strive for energy flexible production planning and control. Beier et al. [16] present a detailed literature review of related research by dividing the relevant energy flexible research approaches into planning and real-time execution. Whereas the planning approaches include organizational methods, the real-time execution targets technical energy flexibility. Relevant technical research approaches are to be found in [13,16-23].

Data communication in energy flexible production systems

The implementation of IoT technology is in progress, thus different levels are covered in literature. The OPC UA interface commonly used in industry can be expanded for energy data transfer. Especially due to the platform-independency, the use of OPC UA is widespread [24]. Bauer et al. [25] abandon the hierarchical automatization pyramid. The concept targeting the adaption of energy demand to supply includes a so-called energy synchronization platform consisting of a market-side and a company-side platform. On the company-side platform the communication model is based on the paradigm *everything as a service*. A factory within this concept is already understood as a cyber-physical production system. Alternative approaches use wireless sensor networks to enable real-time energy monitoring [5,8]. Tan et al. [26] expand the approach of energy monitoring by a benchmark algorithm detecting advanced energetic statuses and conceptually introduce a totally IoT based approach. Shrouf et al. [27] develop an IoT energy management concept based on research, literature and expert interviews including both energy monitoring and a holistic integration of energy data into manufacturing.

3. IoT based closed loop control for energy flexible production systems

The introduced research works according to data communication in smart factories is currently on a conceptual level and not applied to energy flexible control approaches. Therefore, in the following an overall factory concept of an IoT based control loop is introduced and the simulation model structure, the control strategies and the model parametrization are defined and evaluated.

Overall concept of IoT based energy flexible factories

Figure 2 shows the overall factory model for the closed loop control for energy flexible production systems. Both on the component and on the factory level, demand data is measured and communicated to the cloud. Within the cloud, a database includes relevant energy information, e.g. the EII of all components. Furthermore, supply data from on-site generation and the power grid is provided to the cloud. To realize short-term prediction further data could be included, e.g. from weather or energy market forecasts. The implemented control strategy at the cloud computes the control commands from the given information according to the control strategy.

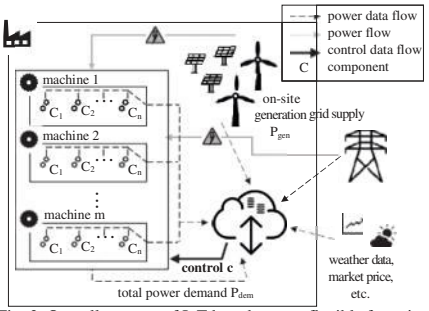


Fig. 2. Overall concept of IoT based energy flexible factories

Simulation model structure

To identify the impact of the closed loop control a simulation model was built up in Matlab Simulink. That model can

ISBN 978-93-88808-61-3

be executed locally or on the IoT platform ThingSpeak in an extended version. The modelling assumptions and simplifications were defined to detect relevant information only. The components¶ behavior is simulated with the following different modules (figure 3).

Functional storage module: Each component is modeled with a so-called functional storage, which is (un)loaded during the component¶V DFWLYH (passive) state. Based on the mean state time of the component, the storage size and the filling (emptying) gradient can be determined. The *internal* control switches the component to active (passive), when the storage reaches the bottom (top) dead center SOC_{bottom} (SOC_{top}).

Convergence module: This module balances the compoQHQW¶V VWDte of charge (SOC)[1] at the end of the simulation to the start value (50 %) to avoid faults during the evaluation.

Reference component module: As a reference component module, a one-to-one copy of the introduced modules only with *internal* control was used to determine the differences between the only internal (storage-based) and the externally controlled (cloud-based) component.

To avoid inefficient control commands and high frequency switching, the external control is allowed in the following SOC range: $SOC \in \{10;90\}\ in\ \%$

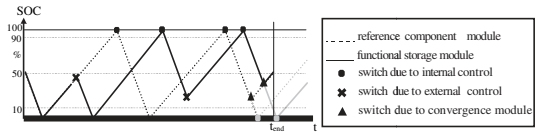


Fig. 3. 62& ¶LPH FRXUVH ZLWK VLPXODWLRQ PRGXOHV¶ IXQFWLRQV

In addition to the component subsystem, the model includes a determination subsystem, which computes the relevant key figures based on the input parameters

- x mean power demand in active state $p_{dem,a}$,
- x mean power demand in passive state $p_{dem,p}$,
- x component or machine status s
- x and the absolute SOC_{abs}.

The resulting key figures for the different control strategies are defined within formulas (1) to (3).

$$|\Delta p| \quad |p_{dem,a} - p_{dem,p}| \tag{1}$$

$$sign(\Delta p) \quad \begin{matrix} -,s\ active \\ +,s\ passive \end{matrix} \tag{2}$$

$$SOC \quad \frac{SOC}{SOC_{top} - SOC_{bottom}^{abs}} \tag{3}$$

To evaluate the impact of the developed energy flexible control strategies, present data were considered, whereas forecasts were neglected initially.

Control Strategies

To adapt the energy demand to the supply, three different control strategies are developed. All considered control strategies are based on the total power demand data (P_{dem}) and on-site generation data (P_{gen}). The difference between the two parameters is defined as the residual power (P_{res}), which is used to describe the interaction of the factory and the power grid (formula 4).

$$\begin{aligned} &P_{res} > 0,\ if\ P_{dem} > P_{gen} \text{ Æ } grid\ supply \\ &P_{res} < 0,\ if\ P_{dem} < P_{gen} \text{ Æ } grid\ feed\text{-}in \\ &P_{res} = 0,\ if\ P_{dem} = P_{gen} \text{ Æ } grid\ neutrality \end{aligned} \tag{4}$$

As third input, component data indices were used, whose specifications depend on the specific control strategy.

Strategy 1: power difference: The simplest decision rule is based on the FRPSRQHQW¶V PHDQ SRZHU GLIIHUHQFH $|\Delta p|$. The mean power differences of all regarded components are sorted by sign and by value. At first, all components with a mean power difference with the same sign as the residual power are excluded. Secondly, the largest remaining power difference is selected and the related component is switched (c = 1). Figure 4 shows the control strategy starting with the component with the maximum value of mean power difference.

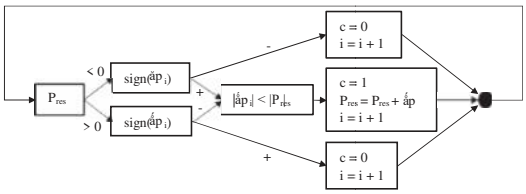


Fig. 4. Scheme of CS 1 (start: max ¸¨S_) DQG &6 2 (start: max/min SOC)

Strategy 2: state of charge: The SOC-control strategy follows the same scheme as strategy 1 (figure 5), but differs in iteration order. Whereas strategy 1 starts with the component i with the largest value of mean power difference, strategy 2 starts with the component holding the smallest/largest SOC.

Strategy 3: best fit: The third control strategy takes an additional static database into account, which includes all possible configurations of the system. For an exemplary five-component-system the corresponding database with all possible current states (rows) and all possible target states (columns) is computed resulting in a matrix with the dimension 2^5 x 2^5, since each component has two different states (active and passive). The matrix contains Δp between one current and one target state. Based on the value of the residual power the best fitting Δp is chosen to determine the target state. The method of control strategy 3 is shown in figure 5.

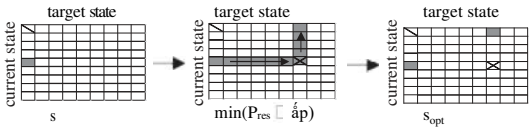


Fig. 5. Scheme of control strategy 3

Model implementation

The simulation model represents a virtual factory consisting of machines of three different types M_1 (4x), M_2 (2x)

[1] SOC = state of charge of the functional storage: term is used by extension to describe storages in general, e.g. electrical storages (battery back), thermal storages, pressure accumulators etc.

and M_3 (4x) and their energy independent components C_{11}, C_{12} (both M_1), C_2 (M_2) and C_3 (M_3). Measured power data of those components are provided in the component model. To consider all machine components, the total power demand on the factory level is based on measured data of five days and scaled regarding the installed amount of flexible energy. The on-site generation data is based on real measured data of radiation and wind during five days in November and scaled by the installed renewable power in the model.

4. Simulation procedure and method evaluation

The simulation model was used with different parametrization to analyze and evaluate the closed loop control considering three different purposes, explained in the following.

Simulation parametrization

Selecting the influencing parameters and configurations, the simulation model should lead to the identification of

- the performance of the control loop regarding the control strategies, the simulation step size and the delay time,
- the most suitable factory configuration considering the amount of energy flexible components and the dimensioning of installed on-site generation and
- the impact of the IoT environment.

Therefore, the simulation model ran according to the parameters shown in table 1.

Table 1. Simulation parameters

Parameter	Characteristics
database	average of a five-day-measurement of demand and on-site generation data
control strategy	CS 1, CS 2, CS 3, CS 12 (□å· CS 1, □å· CS 2), CS 23, CS 13, CS 123 ($^1/_3$ each)
step size	0.1 s, 0.5 s, 1 s
delay time	0.1 s, 15 s, 60 s
energy flexibility	10 %, 18 %, 25 %
Dimensioning of the on-site generation	1:0.5; 1:0.75; 1:1, 1:1.25, 1:1.25, 1:1.5, 1:2
model execution system	local, IoT

The introduced control strategies were applied individually (e.g. CS 1) or in combination by equal weight (e.g. CS 12). The *step size* is a simulation para-meter considering the size of simulation time steps and can be varied manually in the simulation. To ensure model plausibility the parameter specification for the *step size* was chosen in a certain range. The *delay time* describes the time lag within the system which in general occurs in closed control loops. The values for this parameter were considered regarding the minimal *delay time* (due to the model at least as high as the chosen step size) and expected delays within the IoT simulation (higher, not exact computable delay due to communication interfaces). The amount of flexible energy was initially set to 18 % (common value for machine tools [13,28]) and varied up-/downwards. The dimensioning of the on-site generation (DOG) was realized regarding the amount of energy demand, i.e. in case of 1:0.5 the generated amount of energy of five days is half of the energy demand over the same period.

Definition of key performance indicators (KPI)

To evaluate the closed loop control, three different key performance indicators were defined. The determination of all KPIs is based on the resulting residual power with and without application of the developed control loop.

KPI 1: reduction of CO_2 emissions: KPI 1 determines the impact of the control method regarding CO_2 emissions. Grid supply is weighted with the German CO_2 emission factor of 527 g/kWh (power trade balance) [29], whereas on-site generated power is assumed to be renewable and is therefore emission-free.

KPI 2: additional time of grid neutrality: This KPI evaluates the influence of the closed loop control on the time of grid neutrality, i.e. all simulation time steps with $P_{res} = 0$.

KPI 3: cost reduction: KPI 3 considers the economic evaluation concerning the running costs. Due to the newest development within the EEG legislation towards market-regulated feed-in remunerations and the decreasing production costs of renewable energy, the consumption of own-generated power will get more viable in the future. To weight on-site generation and grid supply power, future prices are used according to scenario B in [28] (table 2).

Table 2. Future scenario for energy price development

	Future Scenario	Unit
	Mean energy price	(grid
supply)	0.16	¼/N:K
Feed-in-rewards	0.06	¼/N:K
Own energy production costs	0.05	¼/N:K

Performance of the closed loop control

The performance was evaluated considering three parameters: *step size*, *delay time* and *control strategy*. To analyze and compare their influences, a sensitivity analysis was carried out. Figure 6 shows the sensitivity of the three KPIs for the *step size* (left) and the *delay time* (right).

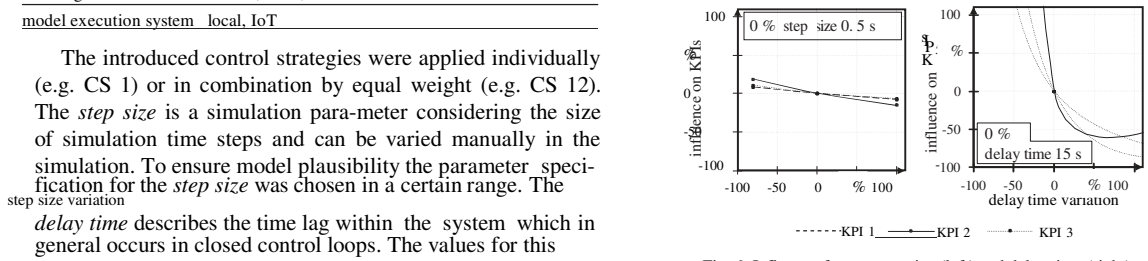


Fig. 6. Influence factors step size (left) and delay time (right)

Both parameters show an inversely proportional influence on the KPIs. Whereas the impact of a changing *step size* is very small, the variation of *delay time* shows a more distinct effect. The sensitivity for *step size* is approximately linear, i.e. in case of further increasing (decreasing) the *step size*, the effect on the KPIs gets equally smaller (higher). The highest sensitivity against the *step size* can be observed for KPI 2

(additional time of grid neutrality). In contrast, the observed impact declines very fast for increasing *delay time*. Nevertheless, a saturation is observed for KPI 2 and KPI 3, which means that delay times higher than a certain threshold do not further decrease the influence. The variation of the *delay time* changes the flexibility of the whole system, and therefore has a significant influence on all KPIs.

The results concerning the *control strategy* are shown in figure 7. The control strategies (x-axis) are sorted by their impact on the KPIs. Control strategy 2 shows the highest influence on all KPIs, whereas the lowest influence is observed for strategy 3. CS 1 is in the same range as CS 2. The insufficient results for control strategy 3 are explicable by the unconsidered input data SOC. In case of external control command in the EORFNHG FRPSRQHQW¶V 62& range, the computed best fit combination of the FRPSRQHQW¶V VWDWHs is not achieved. Both, CS 3 and CS 1, do not consider SOC as decision value. Nevertheless, the impact for strategy 3 is higher due to sequential formation of control commands. Executed simulations with combination CS23 and CS123 result in between the individual control strategies and are neglected in the presentation to ensure clear presentation.

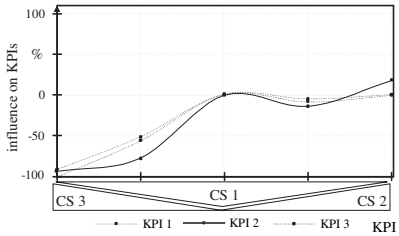


Fig. 7. Influence factor control strategy

To compare the parameters, the average of all KPIs was used to identify differences. Figure 8 shows the *delay time* with the largest impact on the KPIs.

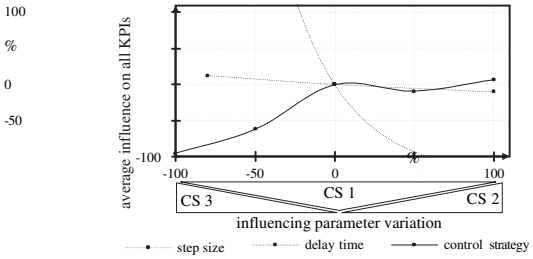


Fig. 8. Comparison of influence parameters

The *control strategy* cannot be treated as a continuous parameter. Comparing control strategy 1 and 2 the impact on the KPIs is as high as the influence of the *step size*. Strategy 3 shows an effect in the range of the *delay time*¶V LQI0XHQFH.

4.4. Factory configuration

The factory configuration was analyzed by regarding the parameters energy flexibility and dimensioning of on-site generation (DOG) as shown in figure 9. The impact of the energy flexibility is rising with increasing amount of flexible energy regarding KPI 1 and KPI 3. The influence of the energy flexibility on KPI 2 (additional time of grid neutrality) is very low in comparison. This can be explained by the variation of *energy flexibility* just based on the flexible energy and neglecting the flexible time of use, i.e. the period, the flexible energy is available. Therefore, increasing (decreasing) *energy flexibility* does not affect time parameters. Concerning the influence of *DOG*, the same effect was observable. Its impact on KPI 2 is lower than on KPI 1 or KPI 3, due to the dimensioning according to the energy amount only. The impact on the reduction of costs (KPI 3) shows a maximum at the dimensioning of 1:1.15. Based on the determination of the cost reduction a maximum close to a 1:1 was expected. The influence on KPI 1 (reduction of CO_2 emissions) is increasing with rising on-site generation. The differences in impact on KPI 1 and KPI 3 can be attributed to differences in their determination. Whereas KPI 1 (reduction of CO_2 emissions) weights grid-supplied power only, the determination of KPI 3 (reduction of costs) includes grid and self-supplied power.

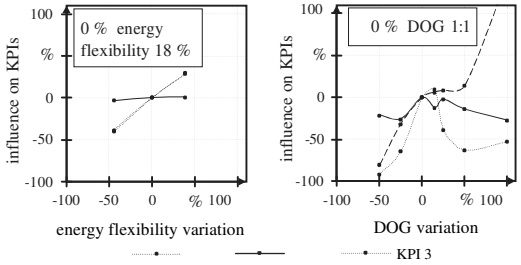


Fig. 9. Influence factors energy flexibility (left) and DOG (right)

4.5. IoT environment

To analyze the impact of the IoT environment, the local model was modified and partly implemented at the cloud. One of the main improvements of cloud-based closed loop control is the centralized accumulation of flexibility information, which is significant for decision making and developing control strategies for all different flexible components in a system of production machines to exploit all given energy flexibility potentials in an optimized way. The cloud-based simulation was executed with the following parameters:

control strategy 1, step size 0.1 s, delay time 0.1 s, energy flexibility 18 %, DOG 1:1. Figure 10 shows the result range of the locally conducted simulations and the IoT result range.

The results of the IoT simulation show conformity with the ORFDO VL**P**X0DWLRQ¶V UHVX0W UH**J**DUGLQ**J** KPI 1 and KPI 3. The IoT model outcome is approximately located in the middle for KPI 3, whereas the results for KPI 1 are in the lower edge. In case of KPI 2 the IoT simulation results do not reach the local simulation results. In IoT-based simulation the occurring delay times are higher than in locally execution and results in less sufficient performance regarding the KPIs. Nevertheless, the developed IoT model is applicable for the desired use case. Further analyses are in progress.

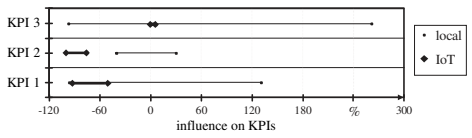


Fig. 10. IoT simulation results versus local simulation results

5. Conclusion and Outlook

The performance analysis indicates that the influence of the parameters differs. It is possible to deduce certain requirements for data communication in general. The delay time has a major influence on all considered KPIs. Therefore, it is important to provide data very fast, whereas the topicality of the data is less important. The results show that data conduction plays a significant role compared to data computing. The control strategies are able to reduce costs and CO_2 emissions and increase the time of grid neutrality. Nevertheless, control strategy 3 shows weak results compared to control strategy 1 or 2. The factory configuration has a higher input than the regarded influencing parameters. Therefore, it is important to implement energy flexibility and on-site generation in early planning steps and apply the closed loop control in addition to ensure most sufficient results. In addition to the conducted simulations further analyzes will be carried out with the IoT model to detect barriers and advantages of the cloud environment. Furthermore, the introduced IoT control loop will be integrated into machine tools to analyze the behavior under real conditions. An IoT communication system is already implemented and will be completed with the closed loop control for flexible production machines.

Acknowledgements

The authors would like to thank the German Federal Ministry of Education and Research (BMBF) and the Project Management Jülich (PtJ) for funding the project *SynErgie* (03SFK3E1).

References

[1] Umweltbundesamt. Erneuerbare Energien in Dtl. - Daten zur Entwicklung im Jahr 2016. Dessau-Roßlau/Germany, 2017. ISSN 2363-829X.

[2] Bundesministerium für Umwelt, Naturschutz, Bau und Reaktorsicherheit (BMUB). Klimaschutzplan 2050 ± Klimaschutzpolitische Grundsätze und Ziele der Bundesregierung. Berlin/Germany, 2016.

[3] Salonitis K, Ball P. Energy efficient manufacturing from machine tools to manufacturing systems. Procedia CIRP 7 (2013), pp. 634±639.

[4] Stoldt J, Franz E, Schlegel A, Putz M. Resource networks: Decentralised factory operation utilising renewable energy sources. Procedia CIRP 26 (2015), pp. 486±491.

[5] Mourtzis D, Vlachou E, Milas N, Dimitrakopoulos G. Energy consumption estimation for machining processes based on real-time shop floor monitoring via wireless sensor networks. Procedia CIRP 57 (2016), pp. 637±642.

[6] Tao F, Zuo Y, Xu LD, Lv L, Lin Z. Internet of Things and BOM-based Life Cycle assessment of energy-saving and emission-reduction of products. IEEE Transactions on Industrial Informatics 10 (2014) Vol. 2, pp. 1252±1261.

[7] Xu W, Yao B, Fang V, Xu W, Liu Q, Zhou Z. Service-oriented sustainable manufacturing: Framework and methodologies. International Conference on Innovative Design and Manufacturing. August 13rd ± 15th, 2014, Montreal, Quebec/Canada. pp. 305±310.

[8] Lu X, Wang S, Li W, Jiang P, Zhang C. Development of a WSN based real time energy monitoring platform for industrial applications. Proceedings of the 2015 IEEE 19th International Conference on Computer Supported Cooperative Work in Design (CSCWD). May 6th ± 8th, 2015, Calabria/Italy. pp. 337±342.

[9] Shrouf F, Ordieres J, Miragliotta G. Smart factories in Industry 4.0: A review of the concept of energy management approached in production based on the Internet of Things Paradigm. Proceedings of the 2014 IEEE IEEM. December 9th ± 12th, 2014, Selangor/Malaysia. pp. 697±700.

[10] Kowalewski S, Bettenhausen KD. Thesen und Handlungsfelder ± Cyber-physical Systems: Chancen und Nutzen aus Sicht der Automation. VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik, 2013.

[11] Reinhart G, Reinhardt S, Graßl M. Energieflexible Produktionssysteme: Einführung zur Bewertung der Energieeffizienz von Produktionssystemen. Werkstatttechnik online 102 Vol. 9 (2014),pp. 622±628.

[12] Zäh MF, Fischbach CWP, Kunkel F. Energieflexibilität in der Produktion identifizieren: Maßnahmen zur Nutzung und Beurteilungsgrößen. ZWF Zeitschrift für wirtschaftlichen Fabrikbetrieb 108 (2013) Vol. 9, pp. 639±642.

[13] Popp RSH, Zaeh MF. Determination of the technical energy flexibility of production systems. Advanced Materials Research 1018 (2014), pp. 365±372.

[14] Popp RSH, Liebl C, Zaeh MF. A multi-level procedure to evaluate the energy flexibility potential of production machines. Procedia CIRP (2017), accepted paper.

[15] Simon P, Schultz C, Keller F, Glasschröder J, Reinhart G. Energieflexibilität in Produktionssystemen. 14. Symposium Energieinnovation. February 10th ± 12th, 2016, Graz/Austria.

[16] Beier J. Simulation approach towards energy flexible manufacturing systems. Cham/Switzerland: Springer, 2017.ISBN 978-3-319-46639-2.

[17] Daryanian B, Bohn RE, Tabors RD. Optimal demand-side response to electricity spot prices for storage-type customers. IEEE Power Engineering Review 9 Vol. 8 (1989), pp. 897±903.

[18] Wang X, Ding H, Qiu M, Dong J. A low-carbon production scheduling system considering renewable energy. Proceedings of IEEE International Conference on Service Operations, Logistics, and Informatics (SOLI). July 10th ± 12th, 2011, Bejing/China. pp. 101±106.

[19] Li L, Sun Z, Yang H, Gu F. Simulation-based energy efficiency improvement for sustainable manufacturing systems. Proceedings of the ASME 2012 International Manufacturing Science and Engineering Conference (MSEC). June 4th ± 8th, 2012, Notre Dame/USA. pp. 1033±1039.

[20] Li L, Sun Z, Tang Z. Real time electricity demand response for sustainable manufacturing systems: challenges and a case Study. 8th IEEE International Conference on Automation Science and Engineering (CASE). August 20th ± 24th, 2012, Seoul/Korea. pp. 353±357.

[21] Zhou Z, Li L. Real time electricity demand response for sustainable manufacturing systems considering throughput bottleneck detection. IEEE International Conference on Automation Science and Engineering (CASE). August 17th ± 20th, 2013, Madison/USA. pp. 640±644.

[22] Sun Z, Li L. Potential capability estimation for real time electricity demand response of sustainable manufacturing systems using Markov decision process. Journal of Cleaner Production 65 (2014), pp. 184±193.

[23] Schultz C, Sellmaier P, Reinhart G. An approach for energy-oriented production control using energy flexibility. Procedia CIRP 29 (2015), pp. 197±202.

[24] Osterfeld H, Klimm B, Langer S, Schultz C, Reinhart G. Mit OPC UA zur energieorientierten Produktionssteuerung. productivity 20 (2015), pp. 49±52.

[25] Bauer D, Abele E, Ahrens R, Bauernhansl T, Fridgen G, Jarke M, Keller F, Keller R, Pullmann J, Reiners R, Reinhart G, Schel D, Schöpf M, Schraml P, Simon P. Flexible IT-platform to synchronize energy demands with volatile markets. Procedia CIRP 63 (2017), pp. 318±323.

[26] Tan YS, Ng YT, Low JSC. Internet-of-Things enabled real-time monitoring of energy efficiency on manufacturing shop floors. Procedia CIRP 61 (2017), pp. 376±381.

[27] Shrouf F, Miragliotta G. Energy management based on Internet of Things: Practices and framework for adoption in production management. Journal of Cleaner Production 100 (2015), pp. 235±246.

[28] Popp RSH, Liebl C, Zaeh MF. Evaluation of the energy flexible operation of machine tool components. Procedia CIRP 63 (2017), pp. 76±81.

[29] Umweltbundesamt. Entwicklung der spezifischen Kohlendioxid-Emissionen des deutschen Strommix in den Jahren 1990 ± 2016. Climate Change (15). Dessau-Rosslau/Germany, 2017

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

A survey of cyber security operations based on Machine learning & Deep learning

k.venkateswarlu
computer science and engineering
MallareddycollegeofEngineering
bagivenky@gmail.com

J.Avinash
computer science and engineering
MallareddycollegeofEngineering
avinashnayakjadhav@gmail.com

***Abstract*— In past decade machine learning (ML) and deep learning (DL), has generated irresistible research interest and attracted unprecedented public attention. With the increasing integration of the Internet and social life, there is change in how people learn and work, but it also exposes them to serious security threats. It is a challenging task to protect sensitive information, data, network and computers connected systems from the unauthorized cyberattacks. For this purpose, effective cyber security is required. Recent technologies such as machine learning and deep learning are integrated with cyberattacks to provide solution to this problem. The paper surveys machine learning and deep learning in cyber security also it discusses the challenges and opportunities of using ML / DL and provides suggestions for research directions.**

Keywords- Cyber security, Machine learning, Deep learning, Intrusion detection.

I. INTRODUCTION

Presently system connected by internet, such as the hardware, software & data can be protected from cyberattacks by means of cyber security. Cybersecurity is a set of technologies and processes designed to protect computers, networks, programs and data from attacks and unauthorized access, alteration, or destruction. As threats become more sophisticated the most recent technologies such as Machine learning (ML) and deep learning (DL) are used in the cybersecurity community to leverage security abilities. Nowadays, cyber security is a stimulating issue in the cyber space and it has been depending on computerization of different application domains such as finances, industry, medical, and many other important areas [11]. To identify various network attacks, particularly not previously seen attacks, is a key issue to be solved urgently [1].

This paper deals with previous work in machine learning (ML) and deep learning (DL) methods for cybersecurity applications and some applications of each method in cyber security operations are described. The ML and DL methods covered in this paper are applicable to detect cyber security threats such as hackers and predators, spyware, phishing and network intrusion detection in ML/DL. Thus, great prominence is placed on a thorough description of the ML/DL methods, and references to seminal works for each ML and DL method are provided [1]. And discuss the challenges and opportunities of using ML / DL for cybersecurity.

The rest of the survey is organized as follows:
Section II tells about cyber security, Section III is composed of Machine learning, Section IV contains survey on Deep learning and Section V dedicated to similarities and differences between Machine learning & Deep learning.

II. CYBER SECURITY

Protection of networks, computer connected devices, programs, and data from malicious attacks or unauthorized access using set of technologies is known as cyber security. Cyber security can be commonly referred as information technology security. Information can be sensitive information, or other types of data for which unauthorized access leads to disaster. In the process of synchronizing with new upcoming technologies, security trends and threat intelligence cyber security are at high risk. However, it is essential to protect information and data from cyberattacks, to maintain cyber security.

A. Challengesofcybersecurity

There are many challenges in the field of cyber security. One of the most challenging elements of cybersecurity is the changing nature of security threats. Traditionally protecting the biggest known threats and not protecting systems against less dangerous risks was approach against maintaining cyber security.

Key challenges of cyber security are:

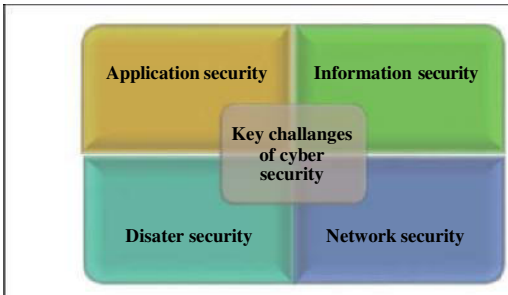


Fig.1. Areas covered in cyber security

- Application security: To protect applications from threats come from faults in the application design, development, deployment, upgrade or maintenance through actions that are taken during the development life-cycle is known as application security. Some basic methods used for application security are:
  1. Input parameter validation.
  2. User/Role Authentication & Authorization.
  3. Session management, parameter manipulation & exception management.
- Information security: It protects information from unauthorized access to save privacy. Methods used are:
  1. Identification, authentication & authorization of user.
  2. Cryptography.

- Disaster recovery planning: It is a process that comprises performing risk assessment, generating priorities, evolving recovery strategies in case of a disaster.
- Network security: Network security includes actions that are used to protect the usability, reliability, integrity and safety of the network. Security componentsinclude:
  1. Anti-virus and anti-spyware.
  2. Firewall, to block unauthorized access to your network.
  3. To identify fast-spreading threats, and Virtual Private Networks (VPNs) and to provide secure remote access intrusion prevention systems (IPS) is needed.

B. Typesofcybersecuritythreats

A cyberattack is a deliberate corruption of computers and servers, electronic systems, networks and data. Cyberattacks uses fake code to alter original computer code, logic or data, resulting in troublemaking consequences that lead to cybercrimes. End goal of cyber security is to prevent cyberattacks.

Following are some common types of cyber threats:

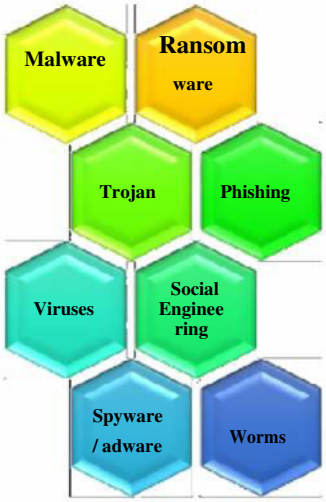


Fig.2. Types of cyber threats

- Type of activity that involves an attacker hacking system files through encryption and demanding a payment to decrypt is known as Ransomware.
- Malware is any file or program used to harm a computer user, such as worms, computer viruses, Trojan horses and spyware.
- Worms are like viruses in that they are self-replicating
- An attack that relies on human interaction to trick users for breaking security to gain sensitive is Social engineering.
- A virus is a piece of malicious code that is loaded onto a machine without the user's knowledge. It spread to other computers by attaching itself to another computerfile.
- Spyware/adware can be installed on computer without knowledge of user when attachments is opened or clicked or downloaded it infects the software and collects personal information.
- Trojan virus is performing malicious activity when executed.
- Phishing is a form of fraud where phishing attacks are sent via email and ask users to click on a link and enter their personal data. However, the intention of these emails is to steal sensitive data, such as credit card or login information. There is a concerning factor about phishing that phishing emails have become sophisticated and often look just like genuine requests for information.

III. MACHINE LEARNING

Machine learning (ML) allows software applications to predict outcomes without being explicitly programmed by use of an algorithm or group of algorithms. The machine learning builds algorithms for receiving input data and uses statistical analysis to predict an output while updating outputs as new data becomes available. Prior work in cyber security based on machine learning and artificial intelligence is presented below.

Liu et al., published a systematic study on security concerns with a variety of machine learning techniques. The existing security attacks explored towards machine learning from two aspects, the training phase and the testing/inferring phase [2]. Furthermore, categorization based on current defensive techniques of machine learning into security assessment mechanisms, countermeasures in the training phase, those in the testing or inferring phase, data security and privacy is done.

Paper presented by Fraley and Dr. Cannady gives better understanding of how machine learning could be leveraged to classify various security events and alerts. They developed model to react to security events by alerting SMEs, alerting analysts or producing reports depending upon the severity of the security event. Additional support for cyber defense was discussed to further reduce the time demand for responding to critical security events [3].

Merat et al. presented different types of computer processes that can be mapped in multitasking environment for the improvement of machine learning. SHOWAN model developed by them was used to learn the cyber awareness behavior of a computer process against multiple concurrent threads [4]. The examined process starts to outperform, and tended to manage numerous tasks poorly, but it gradually learned to acquire and control tasks, in the context of anomaly detection. Finally, SHOWAN plots the abnormal activities of manually projected task and compare with loading trends of other tasks within the group.

In the article, an overview of applying machine learning to address challenges in emerging vehicular networks was presented by Ye et al. This paper introduced basics of machine learning, including major categories and representative algorithms in brief. Some preliminary examples of applying machine learning in vehicular networks to ease data-driven decision making using reinforcement learning was published [5]. Some open issues for further research also highlighted in this paper.

A systematic of the challenges associated with machine learning in the context of big data and categorization based on the V dimensions of big data was published byL'Heureux

r [7]. An overview of ML approaches and how these techniques overcome the various challenges were discussed in this paper. The use of the big data to categorize the challenges of machine learning enables the creation of cause-effect connections for each of the issues. Further, the creation of explicit relations between approaches and challenges enables a more thorough understanding of ML with cyber security.

Golam et al., consider a data-driven next-generation wireless network model, where the MNOs employs advanced data analytics, ML and AI are used for efficient operation, control, and optimization. How ML, AI and computational intelligence play their important roles in data analytics for next-generation wireless networks are discussed in this paper. A set of network designs and optimization schemes with respect to data analytics are presented [8].

Feng and Wu presented a user-centric machine learning system which leverages big data of various security logs, alert information, and analyst insights to the identification of risky user. System provides a complete framework and solution to risky user detection for enterprise security operation center [12]. Generates labels from SOC investigation notes, to correlate IP, host, and users to generate user-centric features, to select machine learning algorithms and evaluate performances, as well as a machine learning system in SOC production environment was briefly introduced. The whole machine learning system is implemented in production environment and fully automated from data acquisition, daily model refreshing, to real time scoring, which greatly improve and enhance enterprise risk detection and management. As to the future work, learning algorithms was proposed for further improvement of the detection accuracy. Technological trends in anomaly detection and identification and open problems and challenges in anomaly detection systems and hybrid intrusion detection systems was discussed by Patcha et al. However, the survey only covers papers published from 2002 to 2006. Unlike Modi C et al., this review covers the application of ML / DL in various areas of intrusion detection and is not limited to cloud security. [1].
Buczak et al. proposed machine-learning methods and their applications to detect intrusion [1]. Algorithms like Neural Networks, Support Vector Machine, Genetic Algorithms, Fuzzy Logics, Bayesian Networks and Decision Tree are also described in paper.

Machine-learning methods are coarsely divided into three major categories as supervised, unsupervised, and reinforcement learning. There are two phases in machine learning i.e. training and testing. In the training stage, a model is learned based on training data, whereas in the testing stage, the trained model is applied to produce the prediction.

A. Supervised Learning

Supervised learning receives a labeled data set and further divide into classification and regression types. Each training sample comes with a discrete (classification) or continuous (regression) value called a label or ground truth. The goal of supervised learning is to gain the mapping from the input feature space to the label or decision space. Classification algorithms assign a categorical label to each incoming sample. Algorithms in this category include Bayesian classifiers, k-nearest neighbors, decision trees, support vector machines, and neural networks [5]. include logistic regression, support vector regression, and
the Gaussian process for regression [3].

B. Unsupervised Learning

For supervised learning, with enough data, the error rate can be reduced close to the minimum error rate bound. However, a large amount of labeled data is often hard to obtain in practice. Therefore, learning with unlabeled data, known as unsupervised learning, has attracted more attention. This method of learning aims to find efficient representation of the data samples, which might be explained by hidden structures or hidden variables, which can be represented and learned by Bayesian learning methods. Clustering is a representative problem of unsupervised learning, grouping samples into different clusters depending on their similarities. Input features could be either the absolute description of each sample or the relative similarities between samples. Classic clustering algorithms include k means, hierarchical clustering, spectrum clustering, and the Dirichlet process. Another important class of unsupervised learning is dimension reduction, which projects samples from a high-dimensional space onto a lower one without losing much information. In many scenarios, the raw data come with high dimension, and may want to reduce the input dimension for various reasons. In optimization, clustering, and classification, the model complexity and the number of required training samples dramatically grow with the feature dimension. Another reason is that the inputs of each dimension are usually correlated, and some dimensions may be corrupted with noise and interference, which will degrade the learning performance significantly if not handled properly

[5]. Some classic dimension reduction algorithms include linear projection methods, such as principal component analysis, and nonlinear projection methods, such as manifold learning, local linear embedding, and isometric mapping[5].

C. Reinforcement Learning

Reinforcement learning deciphers how to map situations to actions, through interacting with the environment in a trial-and-error search to maximize a reward, and it comes without explicit supervision. A Markov decision process (MDP) is generally assumed in reinforcement learning, which introduces actions and (delayed) rewards to the Markov process. The learning Q function is a classic model-free learning approach to solve the MDP problem, without the need for any information about the environment. This Q function estimates the expectation of sum reward when taking an action in a given state, and the optimal Q function is the maximum expected sum reward achievable by choosing actions. Reinforcement learning can be applied in vehicular networks to handle the temporal variation of wireless environments [5].

IV. DEEP LEARNING

Deep Learning is a sub area of Machine Learning research. It is a collection of algorithms in machine learning, used to model high-level abstractions in data. It Uses model architectures composed of multiple nonlinear transformations. Recently, it has made significant advances on various machine-learning tasks. Deep learning aims to understand the data representations, which can be built in supervised, unsupervised, and reinforcement learning. The input layer is at the leftmost, where each node in the figure

mension of input data. The output layer is at the rightmost, corresponding to the desired outputs, whereas the layers in the middle are called hidden layers. Typically, the number of hidden layers and the number of nodes in each layer are. A deep architecture means it has multiple hidden layers in the network as shown in figure 3. However, deeper networks bring new challenges, such as needing much more training data and gradients of networks easily exploding or vanishing. With the help of faster computation resources, new training methods (new activation functions, pretraining), and new structures (batch norm, residual networks), training such deep architecture becomes possible. Deep learning has been widely used in such areas as computer vision, speech recognition, and natural language processing and greatly improved state-of-the-art performance in these areas. Depending on applications, different structures can be added to the deep networks, e.g. convolutional networks share weights among spatial dimensions, whereas recurrent neural networks (RNNs) and long short-term memory (LSTM) share weights among the temporal dimensions [5].

Deep learning aims to learn a hierarchy of features from input data. It can automatically learn features at multiple levels, which makes the system be able to learn complex mapping function directly from data. The most characterizing feature of deep learning is that models have deep architectures. Deep architecture has multiple hidden layers in the network. In contrast a shallow architecture has only a few hidden layers (1 to 2 layers). Deep learning algorithms have been extensively studied in recent years. Algorithms are grouped into two categories based on their architectures:

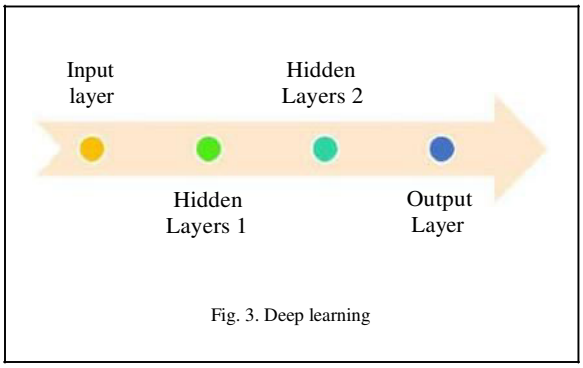


Fig. 3. Deep learning

A. Convolutionalneuralnetworks(CNN)

Convolutional neural networks (CNNs) has gain astonishing recognition in the field of computer vision. It has been continuously advancing the image classification accuracy. Also plays an important role for generic feature extraction such as scene classification, object detection, semantic segmentation, image retrieval, and image caption. Convolutional neural network (CNNs) is most important aspect of deep neural networks in image processing. It is highly effective and commonly used in computer vision applications. The convolution neural network composed of three types of layers: convolution layers, subsampling layers, and full connection layers.

B. RestrictedBoltzmannMachines(RBMs)

RBM is an energy-based probabilistic generative model. It is composed of one layer of visible units and one layer of hidden units. The visible units represent the input vector of a data sample and the hidden units represent features that are abstracted from the visible units. Each visible unit is connected to hidden unit, whereas no connection exists within the visible layer or hidden layer. During past years, the quality of image classification and object detection has been dramatically improved due to the deep learning method.

C. RecurrentneuralNetwork

RNNs are used to make use of sequential information. In a traditional neural network all inputs (and outputs) are independent of each other. To predict the next word in a sentence, need to know which words came before it. RNNs are called recurrent as they perform the same task for every element of a sequence, with the output being depended on the previous computations. RNNs can make use of information in arbitrarily long sequences, but in practice they are limited to only a few steps. An online unsupervised deep learning system is used to filter system log data for analyst. In which variants of Deep Neural Networks (DNNs) and Recurrent Neural Networks (RNNs) are trained to recognize activity of each user on a network and concurrently assess whether user behavior is normal or anomalous, all in real time [10]. Developed model faced several key difficulties in applying machine learning to the cyber security domain. Model was trained continuously in an online fashion, but detection of malicious events was challenging task.

Comparative study was presented by Gavai et al. (2015) of a supervised approach and an unsupervised approach using the isolation forest method for detecting insider threat from network logs. Ryan et al. (1998) applied neural network-based approaches to train network with one hidden layer to predict the probabilities-based network intrusion [10]. A network intrusion was detected for the probability less than

But input features were not structured and did not train the network in an online fashion.

Modeling normal user activity on a network using RNNs was performed by Debar et al. (1992). The RNN was trained on a representative sequence of Unix command line arguments (from login to logout). Network intrusion detected when the trained network poorly predicts the login to logout sequence. While this work partially addresses online training, it does not continuously train the network to consider changing user habits over time.

Recurrent neural networks have been successfully applied to anomaly detection in various alternative domains such as signals from mechanical sensors for machinery such as engines, and vehicles [10].

An inclusive analysis of text Captchas, to evaluate security, a simple, effective and fast attack on text Captchas proposed by Tang et al. Using deep learning techniques, which successfully can attack all Roman character-based text Captchas deployed by the top 50 most popular websites in the world and achieved state-of-the-art results. Success rates range from 10.1% to 90.0% [9]. A novel image-based Captcha named SACaptcha using neural style transfer techniques also presented. This is a positive attempt to

security of Captchas by utilizing deep learning techniques. In this paper, deep learning techniques play two roles: as a character recognition engine to recognize individual characters and as a powerful means to enhance the security of the image-based Captcha This proved that deep learning is a double-edged sword. It can be either used to attack Captchas or improve the security of Captchas [9]. In future, they predicted existing text Captchas are no longer secure. Other Captcha alternatives are robust, and the designs of new Captchas can be simultaneously secure and usable are still challenging difficulties to be work on [9].

A new approach for detection of network intrusion using unsupervised deep learning with iterative K-means clustering proposed by Alom and Taha. In addition, unsupervised ELM, and only K-means clustering approaches were tested. From empirical evaluation on KDD-Cup 99 benchmark, it isobserved that the deep learning approach of RBM and AE with k-means clustering show around 92.12% and 91.86% accuracy for network intrusion detection respectively. RBM with K-means clustering provides around 4.4% and 2.95% better detection accuracy compare to K-means and USELM techniques respectively [11].

Nichols and Robinson present an online unsupervised deep learning approach to detect anomalous network activity from system logs in real time. Models decompose anomaly scores into the contributions of individual user behavior features for increased interpretability to aid analysts reviewing potential cases of insider threat. Using the CERT Insider Threat Dataset v6.2 and threat detection recall, their novel deep and recurrent neural network models outperform Principal Component Analysis, Support Vector Machine and Isolation [10].

V. SIMILARITIES AND DIFFERENCES BETWEEN MACHINE LEARNING & DEEP LEARNING

There are many puzzles about the relationship among ML, DL, and artificial intelligence (AI). Machine-learning is a branch of AI and is closely related to computational statistics, which also focuses on prediction making using computers [1]. whereas DL is a sub-field in machine-learning research. Its motivation lies in the establishment of a neural network that simulates the human brain for analytical learning. It mimics the human brain mechanism to interpret data such as images, sounds and texts [14].

A. Similarities

- Steps involved in ML and DL

ML and DL method primarily uses similar four steps in except feature extraction in DL is automated rather than manual [12].

- Methods used in ML and DL

ML/DL are similar in these three approaches: supervised, unsupervised and semi-supervised. In supervised learning, each instance consists of an input sample and a label. The supervised learning algorithm analyzes the training data and uses the results of the analysis to map new instances. Unsupervised learning that deduces the description of hidden structures from unlabeled data. Because the sample is unlabeled, the accuracy of the algorithm's output cannot be evaluated, and only the key features of the data can be summarized and explained. Semi-supervised learning is a means of combining supervised learning with unsupervised learning. Semi-supervised learning uses unlabeled data when using labeled data for pattern recognition. Using semi-supervised learning can reduce
label efforts while achieving high accuracy [1].

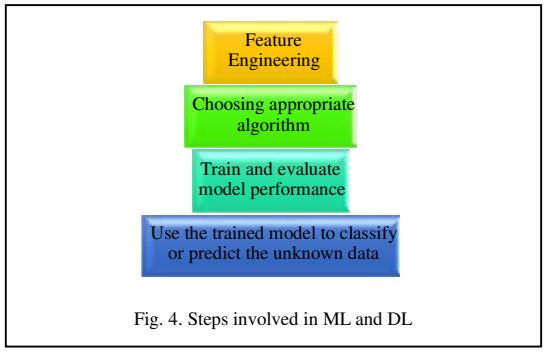


Fig. 4. Steps involved in ML and DL

B. Differences

ML and DL methods different in following ways:

- Data dependencies.

The main difference between deep learning and machine learning is its performance as the amount of data increases. Deep learning algorithms do not perform well when the data volumes are small, because deep learning algorithms require a large amount of data to understand the data perfectly. Conversely, machine-learning algorithm uses the established rules, thus performance is better.

- Hardware dependencies

The DL algorithm requires many matrix operations. The GPU is largely used to optimize matrix operations efficiently. Therefore, the GPU is the hardware necessary for the DL to work properly. DL relies more on high-performance machines with GPUs than machine-learning algorithms.

- Feature processing

The process of putting domain knowledge into a feature extractor to reduce the complexity of the data and generate patterns that make learning algorithms work better is known as feature processing. In ML, most of the characteristics of an application must be determined by an expert and then encoded as a data type. The performance of most ML algorithms depends upon the accuracy of the features extracted. Trying to obtain high-level features directly from data is a major difference between DL and traditional machine-learning algorithms. Thus, DL reduces the effort of designing a feature extractor for each problem.

- Problem-solving method

In Problem-solving method on applying traditional machine-learning algorithms to solve problems, traditional machine learning usually breaks down the problem into multiple sub-problems and solves the sub-problems, ultimately obtaining the result. Unlike deep learning which solves end-to-end problem.

- Execution time.

DL algorithm takes long time to train because there are many parameters in the DL algorithm. Whereas ML training takes relatively less time, only seconds to hours. The test time is exactly opposite for ML and DL. Deep learning algorithms require very little time to run during testing phase compared to ML algorithms. This is not applicable to all ML algorithms, some required short test times [1]

V. CONCLUSION

This paper provides researchers with a strong foundation for making easier and better informed choices about machine learning and deep learning for cyber security. It was reviewed that machine learning has some challenges in handling Big Data whereas deep learning performance is better in context of big data. To improve the security, an innovative image-based captcha named SACaptcha using deep learning techniques can be used. Unsupervised deep learning of RBM and AE with iterative k-means clustering show around 92.12% and 91.86% accuracy for network intrusion detection. In future, system of network intrusion detection for cyber security with online learning approach can be deployed. Machine learning is used to develop a model which detect and highlight advanced malware, by alerting SMEs, alerting analysts or producing reports depending upon the severity of the security event. The model performs these functions with very high accuracy (90%). To detect abnormal network activity from system logs in real time, an online unsupervised deep learning approach can be used that produces interpretable assessments of insider threat in streaming system user logs. This work has therefore accomplished its objective by providing with potential directions for future work and will hopefully serve as groundwork for great improvements of machine learning and deep learning methods for cyber security operations.

REFERENCES

[1] Yang Xin Et Al., "Machine Learning And Deep Learning Methods For Cybersecurity"In IEEE Journals & Magazine, May 2018.

[2] Qiang Liu Et Al., "A Survey On Security Threats And Defensive Techniques Of Machine Learning: A Data Driven View",In IEEE Journals & Magazine, Vol 6, February 2018.

[3] James B. Fraley And Dr. James Cannady, "The Promise Of Machine Learning In Cybersecurity", In Southeast conference , May 2017.

[4] Soorena Merat, P.Eng, Dr. Wahab Almuhtadi, P.Eng., "Artificial Intelligence Application For Improving Cyber-Security Acquirement" In 28th IEEE Canadian Conference On Electrical And Computer Engineering, Halifax, Canada, May 2015.

[5] Hao Ye, Le Liang et al., "Machine Learning for Vehicular Networks" In IEEE Vehicular Technology Magazine, April 2018.

[6] Ge Wang, Jong Chu Ye et al., "Image Reconstruction Is A New Frontier Of Machine Learning", In IEEE Transactions On Medical Imaging ,Vol. 37, pp 1289 – 1296, June 2018.

[7] Alexandra L'heureux et al., "Machine Learning With Big Data: Challenges And Approaches", In IEEE Journal & Magazine ,Vol 5, pp 7776 – 7797, April 2017.

[8] Mirza Golam Kibria Et Al., "Big Data Analytics, Machine Learning And Artificial Intelligence In Next-Generation Wireless Networks", In IEEE Journal & Magazine, May 2018, pp 2169-3536.

[9] Mengyun Tang Et Al., "Research On Deep Learning Techniques In Breaking Text-Based Captchas And Designing Image-Based Captcha", In IEEE Transactions On Information Forensics And Security,Vol13, Issue: 10,pp 2522 – 2537,Oct. 2018.

[10] Aaron Tuor ,Samuel Kaplan And Brian Hutchinson, "Deep Learning For Unsupervised Insider Threat Detection In Structured Cybersecurity Data Streams", In Proceedings Of Ai For Cyber Security Workshop At AAAI ,Dec 2017.

[11] Md Zahangir Alom And Tarek M. Taha, "Network Intrusion Detection For Cyber Security Using Unsupervised Deep Learning Approaches", In IEEE National Aerospace And Electronics Conference (NAECON), Dayton, Oh, USA, June 2017.

[12] Charles Feng*, Shunning Wu And Ningwei Liu, "A User-Centric Machine Learning Framework For Cyber Security Operations Center",
In IEEE International Conference On Intelligence And Security Informatics (ISI), Beijing, China, July 2017.

[13] Ozlem Yavanoglu And Murat Aydos, "A Datasets For Machine Learning Algorithms", In IEEE International
Conference On Big Data ,Jan 2018.

[14] A. Milenkoski, M. Vieira, S. Kounev, A. Avritzer, and B. D. Payne, "Evaluating Computer Intrusion Detection Systems:A Survey of Common Practices," AcmComput.Surv., vol. 48, no. 1, pp. 1–41, 2015.

[15] C. N. Modi and K. Acha, "Virtualization layer security challenges and intrusion detection/prevention systems in cloud computing: a comprehensive review," J. Supercomput., vol. 73, no. 3, pp. 1–43, 2016.

[16] E. Viegas, A. O. Santin, A. França, R. Jasinski, V. A. Pedroni, and L. S. Oliveira, "Towards an Energy-Efficient Anomaly-Based Intrusion Detection Engine for Embedded Systems," IEEE Trans. Comput., vol. 66, no. 1, pp. 163–177, 2017.

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

An Environment Friendly System for Power Saving an Electrical Units by use of Image Processing

1. M.kranthi Assistant Professor Malla Reddy College of Engineering
2. K.Sudha Pavani Assistant Professor Malla Reddy College of Engineering

Abstract: The major problem in the most populated and developing countries like India, is Energy and Power crises. Hence there is a too much need of save energy. We use a several ways to save power like using the electric and electronic gadgets whenever and wherever needed and switching them off while not in use. But there are many places like classrooms, large auditoriums and meeting halls, there will be a fan or an Air-conditioner keeps running in unmanned area too, even before the people arrive. That improves the wastage of power in large amount and contributes to a considerable amount of electricity loss. As we all know about various methods for saving electricity like installing IR sensors to detect people etc. but it is quite costlier and complex in large areas. Here we propose a method of controlling power supply of auditoriums and classrooms using Image Processing. In this firstly we take reference image of empty classroom and if any changes in that reference image accordance with that power supply will turned on and off. This is very simple, efficient and cheaper technique to save energy. Also we include temperature sensor to sense temperature and calculate need of fan or equipments. Another big advantage is, we can extend this project up to application like home automation etc..

Keywords: Picture Processing, Image Partitioning, Edge Detection, Threshold Determination.

I INTRODUCTION

As we all know electricity is basic need of any business and we have to minimize any wastage of electricity. Video surveillance systems are widespread now a day. And it is widely use at airports, banks, casinos and correctional institutions. But now it is increased up to government agencies, business and even schools for increase security and provides video surveillance. As the availability of high speed, broad-band wireless networks and with the proliferation of inexpensive cameras, deploying a large number of cameras for security surveillance has become feasible with economical and technical manner. Several important research questions remain to be addressed before we can rely upon video surveillance as an effective tool for crime prevention, crime resolution and crime protection.

In video surveillance much of the current research focuses on algorithms to analyze video and other media from multiple sources for automatically detecting events. For eg. Intrusion detection, activity monitoring and pedestrian counting. Thus automated power management system is used to detect whether the room is empty or not. By using this technique we monitor changes in the auditorium through sequence of image and accordance with that the power supply is controlled. Image processing is a form of signal processing in that the input is an image and output may be either image or a set of characteristic related to the image.In this implementation first empty image is taken using digital camera then it is converted into gray and by using image enhancement technique we enhanced the image and apply edge detection. In the similar manner real time image is captured, enhanced and edge detected. Now the both images compared to each other and on the basis of results the control signal is generated by using hardware. Both images undergo the following processes

A. Acquisition
B. Gray Conversion
C. Partitioning
D. Edge Detection
E. Comparison
F. Generating control signal.

LITERATURE REVIEW

Literature survey is used to acquire knowledge and skill to complete this project. The main source for gaining the knowledge for this project is latest papers related on this topic. But there are some drawbacks of the previous research, to overcome that drawback and making the project more accurate we are doing several changes for making it more powerful algorithm.

By doing study on the previous research.the following conclusions are taken under consideration.

Accordingly to “Anisha Gupta/Punit Gupta2,jasmeet Chhabra” [3] they proposed intelligent automated system for an efficient power management is being deployed and tested over institutional building in which the lights of the classrooms are automatically controlled by the IOT device. That sense the real time occupancy based on the schedule uploaded on the database server,and takes intelligent action of controlling the lights of classrooms using electromagnet relay switch. The IOT device used here is Intel Galileo board and the sensor used for sensing the real time occupancy in motion detector sensor. The proposed system architectureis explained which include server connected to Intel Galileo board that automatically controls the lights of the class by realizing the real time occupancy of detecting the class using motion sensor[3].

With respect to “N. Sribhagat Verma, Ganesh Taduri ”[4] the need to automate the whole process of power management is very much there and this need is only going to escalate in the future with rising prices and scarcity of resources. Automated power management system is an effort in this direction and a small attempt to solve one of the biggest problems of mankind. With respect to our objective and scope,we have implemented and tested our system to the best possible. thus they calclude that automated power management system provides a practical and feasible approach to the problem of power management.[4]

“Kavya P. Walad, Jyoti Shetty ”[5] they discussed about existing traffic control system and their drawback, to overcome from those drawback can build a flexible traffic light control system based on traffic density. To find traffic density edge detection technique can be used. the edge detection is a well known technique in image processing from identifying an image object, image segmentation, image enhancement. Each edge detection technique have its own advantages and disadvantages in various fields. Gradients based or first order edge detection and Laplace based or second order edge detection operators are discussed in this paper can be implemented in MATLAB. There are so many drawbacks with Gaussian based edge detection is sensitive to noise. This is because of using static dimension of kernel filer and its coefficients. The canny edge detection gives the best performance even in noise condition compared to the first order edge detection. This is more costly compared to the Sobel. Prewitt and Robert’s operator. The main disadvantage with canny is that it has high computational time and responsible for weak edges. The best edge

detection technique is necessary to provide an errorless solution. In future rather than using existing edge detection technique can use fuzzy logic and morphological based edge detection technique for regulating traffic control system based on traffic density to save the time and reduce operating cost.

Accordingly "Manoj Kumar Asst. Professor, Dept of CSE"[6] they calculated all the various steps done and various results are compared with test cases. Students can be at corner or they can be at in front in a group etc. Test case I display two students are sitting and their subtracted image is another image also test case II display two students are sitting and their subtracted image is shown in another image. The study shows that this method is helpful in saving electricity. This method is very cheap, efficient and can reduce wastage of power. This will consistently detect that is there any person in a classroom and auditorium and hence saves electricity.[6] Accordingly "Vankatesh K and Sarath Kumar P "they conclude that image processing is better technique to control the power supply in the auditorium. It shows that it can reduce the wastage of electricity and avoids the free running of those electrical equipments. It is also more consistent in detecting presence of people because it uses real time images. Overall, the system is goog but it still needs improvement to achieve a hundred percent accuracy. If achieved, then we can extend this application to many places like theatres and even for home automation Also they proposed a scope for face detection.

With respect to "Shraddha Dhirde, Priyanka Ghuge, Sneha Khulape " they conclude that monitoring and controlling is done using parameter like temperature and human count by using Rasberry pi3. MB-LBP algorithm is implemented on the attributes of faces of people. This is one of the effective method to control the electric equipment and to reduce power consumption.

Kiiruthika G, Meenatchi R, Mohan raj[9] proposed a system that image processing is one of the useful technique to control the power supply in large areas like malls and auditorium. Also this prevents the free running of electrical application thereby reducing the power wastage. Also it proves to be a consistent and efficient technique to detect the presence of people since it uses real time image.

Patteri Sooraj, Faizankhan Pathan, Gohil Vishal[2] conclude that a classroom can be visualized where all the appliances can be controlled automatically without further human assistance. This makes the camera smart enough to monitor the electrical equipment and thus brings the whole idea of automation into classroom. Hence a lot of efforts and resources can be conserved which can be utilized for different purpose.

Vankatesh K developed a system in that image processing is main keyword to monitor the classroom and control power supply. The drawback with this system is that, it can be used only for the places whose orientation or arrangement is fixed. But they overcome it by resetting the reference images whenever the arrangement is altered. The main program needs not to be altered. Another way of overcoming this limitation is using face detection technique. That is expected to give much flexibility to the overall system.

For overcoming the previous problems related to the work, Here in this recent work we are using same technique of image processing with the temperature sensor and light sensor to sense the atmospheric temperature and light for calculating the need of appliances and making the system more accurate and convenient. In the alternation of face detection we are calculating the centred of object and on the basis of results, the operation will perform through microcontroller programming. In that we firstly take a reference image of empty classroom. This reference image compared to real time image after every 10 seconds. And with respect to changes, the operation will perform. There are many steps and parameter involves in this project that make it better and accurate than before.

CONCLUSION

The study showed that image processing is better technique to control the power supply in the classroom.This shows that it can minimize the wastage of electricity and avoid the free running of equipments. Also by using real time image we make it more consistent in detecting presence of people. Also by adding temperature and light sensor we make this system more accurate and convenient to use.

REFERENCES

[1] Vankatesh K1 and Sarath Kumar P2. "Automatic Real Time Auditorium Power Supply Control Using Image Processing". DOI:03.LSCS.2013.

[2] Patteri Sooraj1, Faizankhan Pthan2, Gohil Vishal3, Pritesh Kukadia4,"Automatic Controlling Of Electrical Appliances in Classroom Using Image Processing".IJERD:April 2016.

[3] Anisha Gupta/Punit Gupta2, Jasmeet Chhabra3."IoT based Power Efficient System Design Using Automation For Classroom". IEEE 2015.

[4] N. SribhagatVarma, Ganesh Taduri, N. BhagirathSai."Automatic Electrical Appliances control Based on Image Processing". IJERT September 2013.

[5] Kavya P Walad,JyothiShetty "Traffic Light Control System Using Image Processing". IJIRCCE October 2014.

[6] ManojKumar."Power Saving in Electrical Devices Using Image Processing". In IJEAST May 2016.

[7] OmkarRamdasGaikwad, Anil Vishwasrao, Prof. KanchanPujari, TejasTalathi "Image Processing Based Traffic Control"in IJSETR April2014.

[8] Shraddha Dhirde1, Priyanka Ghuge2, Sneha Khulape3 "Monitoring and Controlling of an Auditorium " in IJEST 2017.

[9] Kiruthika G1, Meenatchi R1,Mohan Raj V1, Pradeepa S1."Device Power Consumption avoidance using Image Processing" in IJARMET March 2017.

[10] Sunil Kumar Matangiand, SateeshPrathapani, "Design of Smart Power Controlling and Saving System in Auditorium by using MCS 51

[11] VikramadityDangi, AmolParab, KshitijPawar, S.S Rathod,"Image Processing Based Intelligent Traffic Controller", Undergraduate Academic Research Journal(UARJ), ISSN:2278-1129,volume-1, Issue-1,2012.

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

Sixth sense technology: Comparisons and future predictiONS

Sushmitha Chigiri*,Jhansi Rani Marla**
Department of Computer Science & Engineering,
MALLA REDDY COLLEGE OF ENGINEERING
(JNTU),
Hyderabad.

Abstract:
Sixth sense technology is a wearable gestural interface that enhances the physical world around us with the digital information and lets us use natural hand gestures to interact with that information.Sixth sense technology has integrated the real world objects with digital world. it associates technologies like hand gesture reorganization, image capturing, processing, and manipulations etc.

Key words:
sixth sense, hardware components, applications, advantages of sixth sense technology, future predictions in sixth sense technology.

Introduction:
The sixth sense technology is a mini projector coupled with a camera and a cell phone –which acts as the computer and connect to the cloud. All the information is stored on the web. Sixth sense is a new and interesting type of technology which is very easy to use by all the people. Sixth sense is a wearable gestural interface that augments physical word around us with the digital information. Sixth sense technology is a magic where each and everyone can use anywhere we want. It is a gateway between digital and real world.

Evolution:
Sixth Sense was developed at MIT Media Lab by Steve Mann in 1994 and 1997 (head worn gestural interface), and 1998 (neck worn version), and further developed by Pranav Mistry (also at MIT Media Lab), in 2009, both of whom developed both hardware and software for both head worn and neck worn versions of it. It comprises a head worn or neck-worn pendant that contains both a data projector and camera.

Sixth sense:
Every one of us are aware of five basic senses namely seeing, feeling, smelling, tasting, hearing. These senses have evolved millions of years ago .when these senses are not able to do any type of thing then sixth sense came it is just depend on our thinking. In this the information is stored on a paper or a digital storage device. Sixth sense device consists of projector, camera, mirror, web enabled phone and colored markers which are used to track our hand gestures by user hands.

Sixth sense device components

1. Camera
A camera is acting as a digital eye, which sees everything the user sees. The camera is meant to capture and recognize objects in its view and does the tracking of user's hand gestures using techniques based on computer-vision. The camera tracks all the movements made by the thumbs as well as the index fingers of both the hands of the user. On recognizing the object, the camera sends the data to a smart phone for processing.

2. Mobile Component
The sixth sense setup consists of an internet-enabled Smartphone which processes the data send from the camera. Smartphone is used to send and receive data and voice information from anywhere and to anyone through mobile internet. Software is run on the Smartphone which supports this technology and handles data connection. The Smartphone is meant to search the web and to interpret hand gestures. Computer–vision based techniques include programming using Symbian C++ code with more 50,000 lines of code.

3. Projector
The Smartphone interprets the data and this data is projected onto a surface mainly walls, body or hands of a person. A battery is found inside the projector which provides 3 hours battery life. Visual information is projected on to the surfaces and other physical objects which are used as interfaces by the projector. This projection of information is done by a tiny LED projector. The image is projected on to the mirror by the downward facing projector. On touching an object, the information related to the same will appear which will look like the information is part of the object.

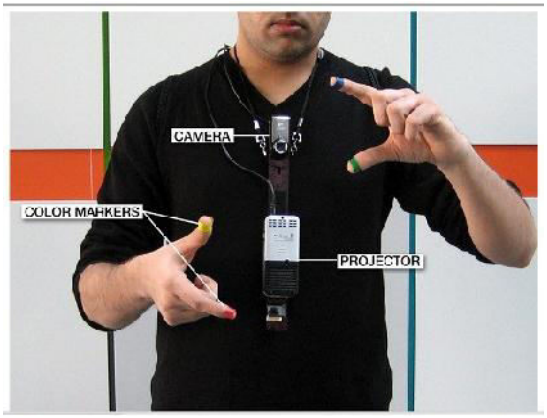
4. Mirror
The mirror is used as the projector hangs from the neck pointing downwards and it reflects the image to a desired surface. This step finally frees the

ISBN 978-93-88808-61-3

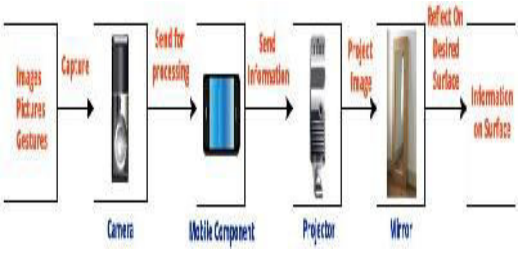
digital information from its confines and places it to the physical world.

5. Color Markers

The color markers that are red, green, blue and yellow are placed at the tips of the fingers which helps the camera to recognize the hand gestures. The various movements and structural arrangements made by these markers are interpreted as gestures that subsequently act as an instruction for the application interfaces are projected.



Working



- The sixth sense implementation hardware is a pendant like wearable mobile interface.
- It consists of a camera, Bluetooth or 3G or Wi-Fi enabled mobile component, projector, mirror and color markers.

Applications: The applications of sixth sense technology are so wide. As already stated, devices with this technology is meant to bring the digital information to the physical real world thereby bridging the existing gap. The recent sixth sense prototype device has showed off its usefulness, flexibility and viability of this technology. The only limitation to this technology is human imagination. Some of the practical applications of this technology are stated below:

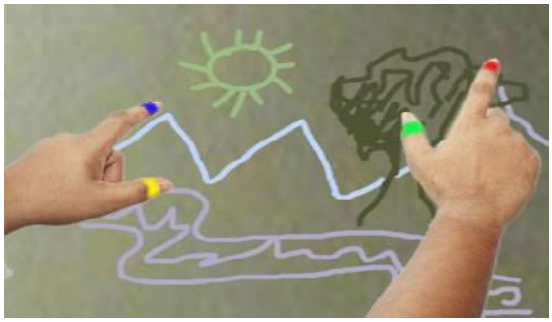
1. Taking Pictures

With the help of framing gestures by hands, user can take pictures of different locations in minimum time. After taking the photos, the user can review the pictures by displaying it onto any surface and then sorting, organizing and resizing the pictures.

2.Viewing Map

Navigation using maps are becoming very common these days. From millions of sources to millions of destinations, this application provides an accurate route along your path. At any instance, Map application helps the user to view any specific location and navigate through it by projecting the map onto a surface. With the help of fingers, mainly the thumb and index fingers, user can zoom in/out or pan the selected area.

3.Drawing Application



This application allows the user to draw on any surface and the drawings are tracked by the movements of finger tips especially the index finger. These pictures can be stored and replaced to any other surface with ease. User with the help of hand gestures can do shuffling through available pictures and drawings.

4.Making Calls

The sixth sense technology supporting device makes calling an easier job. This will project a keypad on your palm or use virtual keypad to make calls thus protecting the privacy. This technique is implemented in other technologies like Skin put. This application helps people with disability to call to a particular number at ease.

5.Interaction with physical objects

The sixth sense technology brings information about different physical objects in minimum time and in a better format. By drawing a circle on the wrist hand, displays an analog watch. Likewise, while reading newspapers, in place of written article, it shows live video news or else, even a paper is capable of providing dynamic information.

6.Grab Information

This technology driven devices are capable of providing information related to any object that is in touch with the user. For example, on holding a book, this device supplies the Amazon or Google ratings of the book as well as the reviews and other relevant details about the book. Also, using this technology, there is no more delay in searching the flight status. The device recognizes the boarding pass and informs the user about the flight status whether it is on time or not.

7.Sixth sense technology does some actions by simple customized hand gestures like an '@' simple when drawn will automatically redirect the user to check mails or a magnifying glass when drawn provides a map onto the surface.

Advantages

- A Sixth sense device has greatest advantage of having a small size and hence it is portable. All the components that make up this device are of light weight and the mobile component fits easily in user's pocket.
- Sixth sense technology comes with an added feature of multi-touch operation and multi-user interaction. Multi-touch operation allows multiple fingers of user to interact with device at a time.
- The prototype is a cost effective device. It consists of components that are common among other devices. The prototype costs around $350 and therefore it sure that when these devices come into the commercial

ISBN 978-93-88808-61-3

market on large scale the cost will become much lower.

- The device allows the user to access digital information at real time from any machine. Brain Computer Interface is not required to access the data.
- Sixth sense technology has broken the limits of a screen onto which a digital data can be projected and manipulated, that is now, data is available on any surface allowing a user to work on the same as per user's convenience.
- The devices are sure to change the habits of computer and machines and make it adaptable to that of humans since hand gestures captured by the device does jobs that was earlier done by machines.
- The software that supports this technology is likely to be an open source code as said by the developer.

Future Additions

As future enhancements, the team is working on to get rid of the use of color markers so as to capture gestures made by hands with ease. Also implementing camera and projector onto a Smartphone or mobile device together can reduce the total space occupied by both and will make it more handy. The team is also working on 3D gesture tracking since nowadays 3D images are a common scenario. This technology can be definitely a fifth sense for a disabled person.

Conclusion

Digital information nowadays is confined to the limits of computer screen or paper. Here, sixth sense technology is taking the natural ways of data display into a new phase which frees data from all its limits and integrates it with the real world seamlessly. It takes out the digital information to the physical world and bridges the gap by bringing the information from the intangible world to the tangible world. Sixth sense technology senses a physical object and projects all information about the object onto any surface let it be wall or hand/ body of a person. This technology is all set to bring that transparent interface to access information about anything and everything around us. Sixth sense technology is definitely the invention of the era and 'Get ready to be part of the magical world.

Reference:

http://www.pranavmistry.com
http://www.media.mit.edu
http://www.wikipedia.org
www.youtube.com
www.google.com
www.ted.com

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

A STATE OF THE ART OF DATABASES TO IMPROVE THE STORAGE EFFICIENCY ON CLOUD COMPUTING

Ch.Vengaiah[1] M.Ghanraj[2]

*T.Vijaykanth Reddy[3], SR.Mahipal[4]

Asst. Professor, Malla Reddy College of Engineering, Hyderabad

Asst. Professor, Malla Reddy College of Engineering, Hyderabad

*Research scholar, Saveetha School of Engineering, Saveetha University

*Research scholar, Hindustan instate of science and technology, Hindustan University

Abstract

The cloud computing is a large groups of remote servers are networked to allow the centralized data storage. It has the access of computer services, resources and can be classified as public, private and reserved. In this study, we explored various types of Data bases used in cloud computing with respect to the category of Knowledge database, XML database, Online databases and Real-Time databases to improve the storage and data efficacy.

Keywords: Cloud computing, Private Cloud, Public Cloud, Knowledge, Real-Time, Bibliographic Database, Bibliographic Database, mobile database.

1. INTRODUCTION

The database in cloud computing is categorized how it interact with various cloud sources for effectively improving the storage capacity for better performances. A data base is a organized collection of data are typically organized to model aspects of reality in a way that supports processes requiring information. For sample, modeling the availability of rooms in hotels in a way that supports finding a hotel with vacancies. And cloud computing is the computing in which large groups of remote servers are networked to allow the centralized data storing, and connected access to computer services or resources. Clouds can be categorizedby manner of public, reserved.

The data base in cloud computing is a storage architecture local administrator to cloud administrator. Traditional databases are organized by *fields*, *records*, and *files*. A field is a single piece of information; a record is one complete set of fields; and a file is a collection of records. To access information from a database, *database management system (DBMS)* is used. This is a collection of programs that supports you to enter, organize, and select data in a database.

The DB techniques are fundamental to increase data availability replication and synchronization. DB is divided into three levels such as front-end, middle-ware, and back-end that is built on Amazon Web Services front end and Mobiledevices is middle ware and also extensible markup language and back end cloud platform. It provides services and also high performance database process have seen exponential development in the past, and such growth is expected to quicken in the future.

2. RELATED WORK

A Cloud database management system (CDBMS) is a distributed database that delivers computing as a service instead of a product. It is the sharing of resources, software, and information between multiple devices over a network which is mostly the internet. Applications of database in real-time are Effective processing complex data

and data with set of the references for expression of the relations between them, Building of Internet-shops and distributed information systems, Building of the virtual company office and virtual kiosks, Storage and reproduction of graphic images, video and audio, Creation of WEB-sites, allotted to unlimited opportunities. Cloud applications connect to a database that is being run on the cloud and have varying degrees of efficiency. In this cloud computing the digital library books barrow problem is occur then they are kept application of homomorphic encryption mechanism for the library

The application of homomorphic encryption mechanism for the library of cloud computing Here they use collision data base[1]. The Design of an Adaptive Peer-to-Peer Network it reduce how it means the cloud of servers support thin clients with various types of service like Web pages and databases. Based on cloud computing peer to peer is now getting very popular

Such techniques are fundamental to increase data availability replication to synchronization have shown useful in the broad context of P2P systems and also super-peer collaborative systems Here they are using mobile data base [2].

Here the real time performs can be down that infrastructure only and Cloud-Mobile Computing Based Real-Time. Cloud-Mobile Computing Based Real-Time in this paper only we introduces a private cloud with SaaS service to realize a real-time video/voice over IP (VVoIP) [3] in this paper only we introduces a private cloud with SaaS they are having the huge capital investment.In their own IT infrastructure and also told that open environment where customers can deploy IT service providers may record service information in a service process from a customer and then collectively deduce the customer's private information

IT service providers may record service information in a service process from a customer and then collectively deduce the customer's private information [4].

Commonly here we are using with the help of computer and internet to get information based on cloud computing only we are share the resources only and also using Xml data base.And It is still in its infancy in regards to its Software as a Service (SaaS), Web Services, Utility Computing and Platform asService (PaaS).The location-based services and the abundant usage of smart phones and GPS-enabled devices. This is necessary to go that outsourcing data has grown rapidly over the past few years .cloud storage and cloud computing services has provided a flexible and cost-effective platform for hosting data from business and individuals in this knowledge-based development in the cloud rule engine and service oriented design in graph database has been designed to operate in a current cloud environment Cloud database are responsible for store data in high available form in cloud environment. The migration to one environment to another is difficult in that case cloud database uses to store and retrieve data

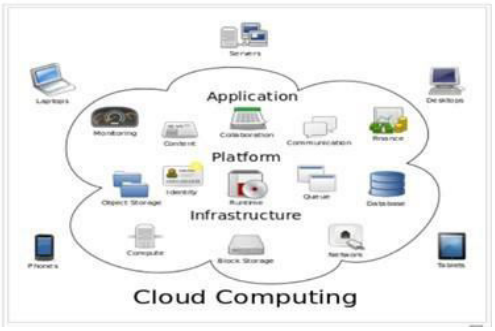


Figure 1. Cloud Computer metaphor: for user, network element

In fig1 cloud computer of metaphor user in network element is interacted with a applications, Infrastructure, and platform with different devices will be used. The privacy preserving system store data base of storage architecture local administrator to cloud administrator for this to learn about the outsourced database content and also more over the machine readable rights expressions are used in order to limit user of the database to a need-to-know basis Here they use cloud data base [5]

ISBN 978-93-88808-61-3

Cloud Storage for Real-Time Databases

Real-time Cloud Storage is a fast and fully managed backend-as-a-service (BaaS) that removes the administrative burden of operating distributed databases while providing seamless scalability. Designed for internet scale applications, Cloud Storage is particularly suited for online collaborative applications due to its powerful real-time notification features. Real time Cloud Storage is the ability of providing real-time notifications when data changes inside the storage. This means that's incredibly easy to develop applications that synch data between several users. Your application simply defines which events are of interest (e.g. table inserts, item updates, item deletes)

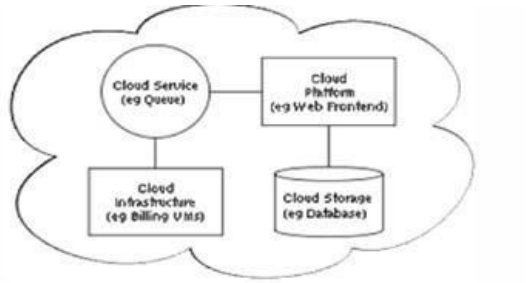


Figure 2.cloud computer sample architecture

Here the real time performs can be down that substructure only. In fig2 cloud computer example architecture A real time system can take advantage of intensive computing capabilities and scalable virtualized environment of cloud computing Here we are using Real-time data base [6] A real-time database is a database system which uses real-time processing to handle workloads whose state is constantly changing.

This differs from traditional databases containing determined data, mostly unaffected by time. For sample, a stock market changes very rapidly and is active. The graphs of the dissimilar marketplaces appear to be very unstable and yet a database has to keep track of current values for all of the markets of the New York Stock Exchange. Real-time processing means that a transaction is processed fast enough for the result to come back and be acted on correct away. Real-time databases are beneficial for accounting, multi-media, process control, and scientific data analysis

Cloud Storage for Knowledge Databases

Knowledge-based development approach for end-user in is a database used in the cloud environment. To practice the knowledge in the cloud rule engine and service oriented design were convoluted. It offers a framework for the user to store the knowledge, facts and actions. Here we use Knowledge database [7].

A Knowledge Database is a store of information that can be searched or browsed using pre-defined classifications. The classifications help to both guide the researcher and understand the context of the information they have found. Knowledge databases don't just leave users with a search box and let them work out what they should be looking for, knowledge databases provide knowledge database is a technology used to store complex structured and unstructured information used by a computer system.

The original use of the term knowledge-base was to describe one of the two sub-systems of a knowledge-based system. A knowledge-based system consists of a knowledge-base that represents facts about the world and an inference engine that can reason about those facts and use rules and other forms of logic to deduce new facts or highlight inconsistencies

- Flat data. Data was usually represented in a tabular format with strings or number in each field.
- Multiple users. A conventional database must support more than one user or system logged into the same data at the same time.
- Transactions. An essential requirement for a database was to maintain integrity and consistency among data that is accessed by concurrent users. These are the so-called ACID properties

Cloud Storage for XML Databases

Here using with the help of computer and internet to get information based on cloud computing only we are share the resources only and also using Xml data base .And It is still in its infancy in regards to its Software as a Service (SaaS), Web Services, Utility Computing and Platform as Service (PaaS) Here we use Xml database[8]. An **XML database** is a data persistence software system that allows data to be stored in XML format. These data can then be queried, exported and serialized into the desired format. XML databases are usually associated with document-oriented databases

- **XML-enabled**: these may either map XML to traditional database structures (such as a relational database[2]), accepting XML as input and rendering XML as output, or more recently support native XML types within the traditional database. This term implies that the database processes the XML itself (as opposed to relying on middleware).
- **Native XML** (NXD): the internal model of such databases depends on XML and uses XML documents as the fundamental unit of storage, which are, however, not necessarily stored in the form of text files.

XML in databases: the increasingly common use of XML for data transport, which has meant that "data is extracted from databases and put into XML documents and vice-versa".It may prove more efficient (in terms of conversion costs) and easier to store the data in XML format. In content-based applications, the ability of the native XML database also minimizes the need for extraction or entry of metadata to support searching and navigation. In a native XML environment, the entire content store becomes metadata through query languages such as X Path and XQuery, including content, attributes and relationships within the XML.

Cloud Storage for On-Line Databases In high performance database process databases have seen exponential growth in the past, and such growth is expected to accelerate in the future to increases the storage capacity comparing to old to implement like new thing Here we use the online database [9].

An online database is a database accessible from a network, including from the Internet.

It differs from a local database, held in an individual computer or its attached storage, such as a CD.

- For the system or software designed to Currently, there are several database products designed specifically as hosted databases delivered as software as a service products. These differ from typical traditional databases such as Oracle, Microsoft SQL Server, Sybase, etc. Some of the differences are:
- These online databases are delivered primarily via a web browser
- They are often purchased by a monthly subscription
- They embed common collaboration features such as sharing, email notifications, etc.

Cloud Storage for Bibliographic Database

In cloud computing research and selection system they are using the out ranking method because to get a better refine the results and also main contribution is conceiving an Agent that uses both the Skyline. Here we use Bibliographic databases[10].

The database of bibliographic records, an organized digital collection of references to published literature, including journal and newspaper articles, conference proceedings, reports, government and legal publications, patents, books, etc. In contrast to library catalogue entries, a large proportion of the bibliographic records in bibliographic databases describe articles, conference papers, etc., rather than complete monographs, and they generally contain very rich subject descriptions in the form of keywords, subject classification terms, or abstracts.

A bibliographic database may be general in scope or cover a specific academic discipline. A significant number of bibliographic databases are still proprietary, available by licensing agreement from vendors, or directly from the indexing and abstracting services that create them. Many bibliographic databases evolve into digital libraries, providing the full-text of the indexed contents. Others converge with non-bibliographic scholarly databases to create more complete disciplinary search engine systems, such as Chemical Abstracts.

Cloud Storage for mobile database

Here Designing and developing we use the three levels front-end, middle-ware, and a back-end that is built on Amazon Web Services front end is Mobil devise and middle ware is extensible markup language and back end cloud platform provides services Here we use theRelational Database [11].

A **mobile database** is either a stationary database that can be connected to by a mobile computing device (e.g., smartphones and PDAs) over a mobile network, or a database which is actually stored by the mobile device. This could be a list of contacts, price information, distance travelled, or any other information.[1]

Many applications require the ability to download information from an information repository and operate on this information even when out of range or disconnected. An example of this is your contacts and calendar on the phone. In this scenario, a user would require access to update information from files in the home directories on a server or customer records from a database. This type of access and work load generated by such users is different from the traditional workloads seen in client–server systems

Cloud Storage for collision Database

The Design of an Adaptive Peer-to-Peer Network it reduce how it means the cloud of servers support thin clients with various types of service like Web pages and databases. On based on cloud computing peer to peer is now getting very popular. Here we use collision database [12]. Collision induced absorption and emission refers to spectral features generated by inelastic collisions of molecules in a gas. Such inelastic collisions (along with the absorption or emission of photons) may induce quantum transitions in the molecules, or the molecules may form transient supra molecular complexes with spectral features different from the underlying molecules. Collision-induced absorption and emission is particularly important in dense gases, such as hydrogen and helium clouds in found in astronomical systems.

cloud storage for Time-series data base

In this real-time services they are having the huge capital investment in their own IT infrastructure and also told that open environment where customers can deploy IT service

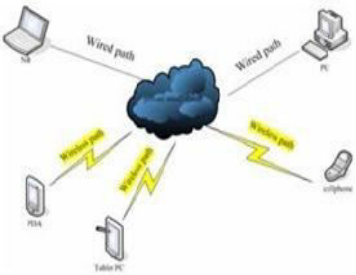


Figure 3.Cloud-mobile Computer via wired or wireless

providers may record service information in fig3 Cloud-mobile Computer via wired or wireless service process from a customer and then collectively deduce the customer's private information Here we use Time-series data base.

A time series database (TSDB) is a software system that is optimized for handling time series data, arrays of numbers indexed by time (a date time or a date time range). In some fields these *time series* are called profiles, curves, or traces. A time series of stock prices might be called a price curve. A time series of energy

ISBN 978-93-88808-61-3

consumption might be called a load profile. A log of temperature values over time might be called a temperature trace.

Despite the disparate names, many of the same mathematical operations, queries, or database transactions are useful for analysing all of them. The implementation of a database that can correctly, reliably, and efficiently implement these operations must be specialized for time-series data.

cloud storage for Spatial database

The location-based services and the abundant usage of smart phones and GPS-enabled devices. This is necessary to go that outsourcing data has grown rapidly over the past few years .cloud storage andcloud computing services has provided a flexible and cost-effective platform for hosting data from businesses and individuals Here we use Spatial database [13].

A spatial database, or geodatabase is a database that is optimized to store and query data that represents objects defined in a geometric space. Most spatial databases allow representing simple geometric objects such as points, lines and polygons. Some spatial databases handle more complex structures such as 3D objects, topological coverages, linear networks, and TINs. While typical databases are designed to manage various numeric and character types of data, additional functionality needs to be added for databases to process spatial data types efficiently. These are typically called geometry or feature. The Open Geospatial Consortium created the Simple Features specification and sets standards for adding spatial functionality to database systems

cloud storage for graph database

XGDBench is a graph database fig 4 Architecture of XGDB has been designed to operate in a current cloud environment. Cloud service benchmark to the domain of database bench-mark. It emphases on exascale cloud. This bench is centered on MAG model for realistic demonstrating of characteristic graphs Here we use graph database [14].

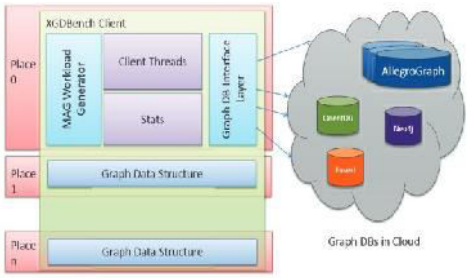


Figure 4. Architecture of XGDB

Graph databases have grown into increasingly popular for a variability of customs ranging from modeling to tracking software engineering enslavements in fig5 Virtual hierarchy as a graph. These extents use graphs because it expresses the awkward in graph traversal. Including migration this is used in hybrid cloud. It will provide a dramatic gain in concert. These databases solve the difficult in cloud management. The graph language database is very dominant. Here we use graph database [15]

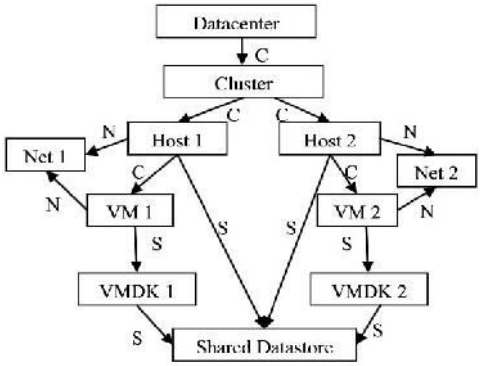


Figure 5. Virtual hierarchy as a graph

Cloud database are responsible for store data in high available form in cloud environment in fig5. The migration to one environment to another is difficult in that case cloud database uses to store and retrieve data. This provides a official way for drifting data among HBase as a column family database to Neo4j as graph databaseHere we use graph database [16].

3. OBSERVATIONS

1. Here we observe that the privacy preserving system store data base of storage architecture local administrator to cloud administrator
2. In this real time database we observe that intensive computer capabilities.
3. A private cloud with SaaS service to a real-time video, voice over IP.
4. To storing data in cloud computing is to get a better refine the result.
5. P2p system is a super-peer collaborative system.
6. In mobile device we use the three layers front end, middleware, backend for Designing and developing
7. The online database process is high performance have seen in exponential growth is past
8. In online database growth will be increases the storage capacity comparing to old.
9. Using the web pages of database we reduce servers support in thin client of peer to peer.

4. CONCLUSION

Finally we conclude the survey of database in cloud computing to improve the storage and data effectively. And here we use Various types of Data base in cloud computing like Bibliographic database, Knowledge database, XML database,Online databases, Real- time databases, Bibliographic Database.

5.REFERENCES

[1] Qingjie MENG, ChangqingGONG,Research of cloud computing security in digital library 6th International Conference on Information Management, Innovation Management and Industrial Engineering 2013.

[2] Barcelona, Spain,Data Replication and Synchronization in P2P Collaborative Systems 26th IEEE International Conference on Advanced Information Networking and Applications 2012.

[3] BaoRong Chang, Hsiu Fen Tsai, Chi-Ming, ChenYi-Sheng Chang, Chien-Feng Huang Cloud-Mobile Computing Based Real- Time VVoIP with PSO-ANFIS Tuning Conference on Technologies and Applications of Artificial Intelligence 2013.

[4] Gaofeng Zhang, Yun Yang, Xiao Liu, JinjunChen,A Time-series Pattern based Noise Generation Strategy for Privacy Protection in Cloud Computing 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing 2012. Computing,11th IEEE International Conference on Computer and Information Technology 2011.

[5] Rui Zhou, Jing Li, Jinghan Wang, GuoweiWang,A Knowledge-Based Development Approach with Fact and Service for End-User in Cloud Computing IEEE 37th Annual Computer Software and Applications Conference Workshops 2013.

[6] TarunKarnwal, T. Sivakuma, G. Aghila A Comber Approach to Protect Cloud Computing against XML DDoS and HTTP DDoS attack IEEE Students' Conference on Electrical, Electronics and Computer Science 2012.

[7] David Taniar,High Performance Database Processing 26th IEEE International Conference on Advanced Information Networking and Applications 2012.

[8] ManarABOUREZQ andAbdellah IDRISSI,Introduction of an outranking method in the cloud computing research and selection system based on the skyline.

[9] MaziyarShariatPanahi, Peter Woods, Peter Wood,Designing and developing a location-based mobile tourism application by using cloud-based platform International Conference on Technology, Informatics, Management, Engineering & Environment

(TIME-E 2013) Bandung, Indonesia, June 23-26,2013 2013.

[10] Apirajitha.P.S., AnithaAngayarkannA Design of an Adaptive Peer-to-Peer Network to Reduce Power Consumption Using Cloud Computing IEEE International Conference on Advanced Communication Control and ComputinTechnologies (ICACCCT) 2012.

[11] Ling Hu, Wei-Shinn Ku, SpiridonBakiras, Cyrus Shahab,Spatial Query Integrity with Voronoi Neighbors IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, no 4, April 2103.

[12] MiyuruDayarathna, ToyotaroSuzumura,XGDBench:

[13] VijayaraghavanSoundararajan and ShishirKakaraddi ,Applying Graph Databases to Cloud Management: An Exploration IEEE International Conference on Cloud Engineering 2014

[14] Mahdi NegahiShirazi, Ho Chin Kuan, HosseinDolatabadi,Design Patterns to Enable Data Portability between Clouds' Databases 12th International Conference on Computational Science and Its Applications 2012.

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

Securing Cloud Data under Key Exposure

Jangam Ravali,

M.Tech. Scholar,

Cse Department,
Malla Reddy College of Engineering

Abstract—Recent news reveal a powerful attacker which breaks data confidentiality by acquiring cryptographic keys, by means of coercion or backdoors in cryptographic software. Once the encryption key is exposed, the only viable measure to preserve data confidentiality is to limit the attacker's access to the ciphertext. This may be achieved, for example, by spreading ciphertext blocks across servers in multiple administrative domains—thus assuming that the adversary cannot compromise all of them. Nevertheless, if data is encrypted with existing schemes, an adversary equipped with the encryption key, can still compromise a single server and decrypt the ciphertext blocks stored therein. In this paper, we study data confidentiality against an adversary which knows the encryption key and has access to a large fraction of the ciphertext blocks. To this end, we propose Bastion, a novel and efficient scheme that guarantees data confidentiality even if the encryption key is leaked
and the adversary has access to almost all ciphertext blocks. We analyze the security of Bastion, and we evaluate its performance by means of a prototype implementation. We also discuss practical insights with respect to the integration of Bastion in commercial dispersed storage systems. Our evaluation results suggest that Bastion is well-suited for integration in existing systems since it incurs less than 5% overhead compared to existing semantically secure encryption modes.

Index Terms—Key exposure, data confidentiality, dispersed storage.

1 INTRODUCTION

THE world recently witnessed a massive surveillance program aimed at breaking users' privacy. Perpetrators were not hindered by the various security measures deployed within the targeted services [31]. For instance, although these services relied on encryp- tion mechanisms to guarantee data confidentiality, the necessary keying material was acquired by means of backdoors, bribe, or coercion.

If the encryption key is exposed, the only viable means to guarantee confidentiality is to limit the ad- versary's access to the ciphertext, e.g., by spreading it across multiple administrative domains, in the hope that the adversary cannot compromise all of them. However, even if the data is encrypted and dispersed across different administrative domains, an adversary equipped with the appropriate keying material can compromise a server in one domain and decrypt cipher- text blocks stored therein.

In this paper, we study data confidentiality against an adversary which knows the encryption key and has access to a large fraction of the ciphertext blocks. The adversary can acquire the key either by exploiting flaws or backdoors in the key-generation software [31], or by compromising the devices that store the keys (e.g., at the user-side or in the cloud). As far as we are aware, this adversary invalidates the security of most cryptographic solutions, including those that protect encryption keys by means of secret-sharing (since these keys can be leaked as soon as they are generated).

To counter such an adversary, we propose Bastion, a novel and efficient scheme which ensures that plaintext data cannot be recovered as long as the adversary has access to at most all but *two* ciphertext blocks, even when the encryption key is exposed. Bastion achieves this by combining the use of standard en- cryption functions with an efficient linear transform. In this sense, Bastion shares similarities with the no- tion of all-or-nothing transform. An AONT is not an encryption by itself, but can be used as a pre-processing step before encrypting the data with a block cipher. This encryption paradigm—called AON encryption— was mainly intended to slow down brute-force attacks on the encryption key. However, AON encryption can also preserve data confidentiality in case the encryp- tion key is exposed, as long as the adversary has ac- cess to at most all but one ciphertext blocks. Existing AON encryption schemes, however, require *at least* two rounds of block cipher encryptions on the data: one pre- processing round to create the AONT, followed by an- other round for the actual encryption. Notice that these rounds are sequential, and cannot be parallelized. This results in considerable—often unacceptable—overhead to encrypt and decrypt large files. On the other hand, Bastion requires only one round of encryption—which makes it well-suited to be integrated in existing dis- persed storage systems.

We evaluate the performance of Bastion in compari- son with a number of existing encryption schemes. Our results show that Bastion only incurs a negligible per-

- *G. Karame is affiliated with NEC Laboratories Europe, Heidelberg, 69115 Germany. E-mail: ghassan.karame@neclab.eu*
- *C. Soriente and S. Capkun are affiliated with the Compute Science Department of ETH Zurich, 8092, Switzerland. Email: first-name.lastname@inf.ethz.ch*
- *K. Lichota is affiliated with 9livesdata, Poland. Email: li- chota@9livesdata.com*

ISBN 978-93-88808-61-3

formance deterioration (less than 5%) when compared to symmetric encryption schemes, and considerably improves the performance of existing AON encryption schemes [12], [26]. We also discuss practical insights with respect to the possible integration of Bastion in commercial dispersed storage systems. Our contribu- tions in this paper can be summarized as follows:

- We propose Bastion, an efficient scheme which ensures data confidentiality against an adversary that knows the encryption key and has access to a large fraction of the ciphertext blocks.
- We analyze the security of Bastion, and we show that it prevents leakage of any plaintext block as long as the adversary has access to the encryption key and to all but two ciphertext blocks.
- We evaluate the performance of Bastion analytically and empirically in comparison to a number of existing encryption techniques. Our results show that Bastion considerably improves (by more than 50%) the performance of existing AON encryption schemes, and only incurs a negligible overhead when compared to existing semantically secure encryption modes (e.g., the CTR encryption mode).
- We discuss practical insights with respect to the deployment of Bastion within existing storage systems, such as the HYDRAstor grid storage system [13], [23].

The remainder of the paper is organized as follows. In Section 2, we define our notation and building blocks. In Section 4, we describe our model and introduce our scheme, Bastion. In Section 5, we analyze our scheme in comparison with a number of existing encryption primitives. In Section 6, we implement and evaluate the performance of Bastion in realistic settings; we also discuss practical insights with respect to the integration of Bastion within existing dispersed storage systems. In Section 7, we overview related work in the area, and we conclude the paper in Section 8.

2 PRELIMINARIES

We adapt the notation of [12] for our settings. We define a block cipher as a map $F : \{0, 1\}^k \times \{0, 1\}^l \rightarrow \{0, 1\}^l$, for positive k and l. If P_l is the space of all $(2^l)!$ l- bits permutations, then for any $a \in \{0,1\}^k$, we have $F(a, \cdot) \in P_l$. We also write $F_a(x)$ to denote $F(a, x)$. We model F as an ideal block cipher, i.e., a block cipher picked atrandom from $BC(k, l)$, where $BC(k, l)$ is the space of all block ciphers with parameters k and l. For a given block cipher $F \in BC(k, l)$, we denote $F^{-1} \in BC(k, l)$ as $F^{-1}(a, y)$ or as $F_a^{-1}(y)$, for $a \in \{0, 1\}^k$.

Encryption modes

An encryption mode based on a block cipher F/F^{-1} is given by a triplet of algorithms $\mathcal{Q} = (\mathsf{K}, \mathsf{E}, \mathsf{D})$ where:

K The key generation algorithm is a proba- bilistic algorithm which takes as input a security parameter k and outputs a key $a \in \{0, 1\}^k$ that specifies F_a and F_a^{-1}.

E The encryption algorithm is a probabilistic algorithm which takes as input a message $x \in \{0, 1\}^*$, and uses F_a and F_a^{-1} as oracles to output ciphertext y.

D The decryption algorithm is a deterministic algorithm which takes as input a ciphertext y, and uses F_a and F^{-1} as oracles to output plaintext $x \in \{0, 1\}^*$, or $\perp$ if y is invalid.

For correctness, we require that for any key $a \leftarrow \mathsf{K}(1^k)$, for any message $x \in \{0, 1\}^*$, and for any $y \leftarrow \mathsf{E}^{F_a, F_a^{-1}}(x)$, we have $x \leftarrow \mathsf{D}^{F_a, F_a^{-1}}(y)$.

Security is defined through the following chosen-plaintext attack (CPA) game adapted for block ciphers:

$\mathbf{Exp}^{ind}(A, b)$
$F \leftarrow BC(k, l)$
$a \leftarrow \mathsf{K}(1^k)$
$x_0\, x_1\, state \leftarrow A^{\mathsf{E}^{F_a, F_a^{-1}}}(find)$
$y_b \leftarrow \mathsf{E}^{F_a, F_a^{-1}}(x_b)$
$b' \leftarrow A(guess, y_b, state)$

In the *ind* experiment, the adversary has unrestricted oracle access to $\mathsf{E}^{F, F^{-1}}$ during the "find" stage. At this point, A outputs two messages of equal length x_0, x_1, and some *state* information that are passed as input when the adversary is initialized for the "guess" stage (e.g., *state* can contain the two messages x_0, x_1). During the "guess" stage, the adversary is given the ciphertext of one message out of x_0, x_1 and must guess which message was actually encrypted. The advantage of the adversary in the *ind* experiment is:

$$\mathrm{Adv}^{ind}_{\mathcal{Q}}(\mathrm{A}) = |\Pr[\mathrm{Exp}^{ind}_{\mathcal{Q}}(A, 0) = 1] - \Pr[\mathrm{Exp}^{ind}_{\mathcal{Q}}(A, 1) = 1]|$$

Definition 1. An encryption mode $\mathcal{Q} = (\mathsf{K}, \mathsf{E}, \mathsf{D})$ is *ind* secure if for any probabilistic polynomial time (p.p.t.) adversary A, we have $\mathrm{Adv}^{ind}_{\mathcal{Q}}(\mathrm{A}) \leq \varrho$, where ϱ is a negligible function in the security parameter.

REMARK 1. The *ind* experiment allows the adversary to see the entire (challenge) ciphertext. In a sce- nario where ciphertext blocks are dispersed across a number of storage servers, this means that the ind- adversary can compromise all storage servers and fetch the data stored therein.

REMARK 2. In the *ind* experiment (and in other experiments used in this paper), we adopt the Shannon Model of a block cipher that, in practice, instanti- ates an independent random permutation for every different key. This model has been used in previous

ISBN 978-93-88808-61-3

related work [3], [12], [17] to disregard the algebraic or cryptanalysis specific to block ciphers and treat them as a black-box transformation.

All or Nothing Transforms

An All or Nothing Transform (AONT) is an efficiently computable transform that maps sequences of input blocks to sequences of output blocks with the following properties: *(i)* given all output blocks, the transform can be efficiently inverted, and *(ii)* given all but one of the output blocks, it is infeasible to compute any of the original input blocks. The formal syntax of an AONT is given by a pair of p.p.t. algorithms $\mathcal{Q} = (\mathsf{E}, \mathsf{D})$ where:

- E The encoding algorithm is a probabilistic algorithm which takes as input a message $x \in \{0,1\}^*$, and outputs a pseudo-ciphertext y.
- D The decoding algorithm is a deterministic algorithm which takes as input a pseudo- ciphertext y, and outputs either a message $x \in \{0,1\}^*$ or $\perp$ to indicate that the input pseudo-ciphertext is invalid.

For correctness, we require that for all $x \in \{0,1\}^*$, and for all $y \leftarrow \mathsf{E}(x)$, we have $x \leftarrow \mathsf{D}(y)$.

The literature comprises a number of security definitions for AONT (e.g., [8], [12], [26]). In this paper, we rely on the definition of [12] which uses the *aont* ex- periment below. This definition specifies a block length l such that the pseudo-ciphertext y can be written as $y = y[1] \ldots y[n]$, where $|y[i]| = l$ and $n \geq 1$.

$\mathbf{Exp}^{aont}(A, b)$
$x, state \leftarrow A(find)$
$y_0 \leftarrow \mathsf{E}(x)$
$y_1 \leftarrow \{0,1\}^{|y_0|}$
$b' \leftarrow A^{Y_b}(guess, state)$

On input j, the oracle Y_b returns $y_b[j]$ and accepts up to $(n-1)$ queries. The *aont* experiment models an adversary which must distinguish between the encod- ing of a message of its choice and a random string (of the same length), while the adversary is allowed access to all but one encoded blocks. The advantage of A in the *aont* experiment is given by:

$$\mathrm{Adv}^{aont}_{\mathcal{Q}}(A) = |\Pr[\mathrm{Exp}^{aont}(A, 0) = 1] - \Pr[\mathrm{Exp}^{aont}(A, 1) = 1]|$$

Definition 2. An All-or-Nothing Transform $\mathcal{Q} = (\mathsf{E}, \mathsf{D})$ is *aont*-secure if for any p.p.t. adversary A, we have $\mathrm{Adv}^{aont}_{\mathcal{Q}}(A) \leq \varrho$, where ϱ is a negligible function in the security parameter.

Known AONTs

Rivest [26] suggested the *package transform* which leverages a block cipher F/F^{-1} and maps m block strings to $n = m + 1$ block strings. The first $n-1$ output blocks are computed by XORing the i-th plaintext block with $F_K(i)$, where K is a random key. The n-th output block is computed XORing K with the encryption of each of the previous output blocks, using a key K_0 that is publicly known. That is, given $x[1] \ldots x[m]$, the package transform outputs $y[1] \ldots y[n]$, with $n = m + 1$, where:

$$y[i] = x[i] \oplus F_K(i),\ 1 \leq i \leq n-1,$$
$$y[n] = K \bigoplus_{i=1}^{n-1} F_{K_0}(y[i] \oplus i).$$

Desai [12] proposed a faster version where the block cipher round which uses K_0 is skipped and the last output block is set to $y[n] = K \bigoplus_{i}^{n-1} y[i]$. Both AONTs are secure according to Definition 2 [12].

Remark 3. Although most proposed AONTs are based on block ciphers [12], [26], an AONT is not an encryption scheme, because there is no secret-key information associated with the transform. Given all the output blocks of the AONT, the input can be recovered without knowledge of any secret.

3 System and Security Model

In this section, we start by detailing the system and security models that we consider in the paper. We then argue that existing security definitions do not capture well the assumption of key exposure, and propose a new security definition that captures this notion.

System Model

We consider a multi-cloud storage system which can leverage a number of commodity cloud providers (e.g., Amazon, Google) with the goal of distributing trust across different administrative domains. This "cloud of clouds" model is receiving increasing attention nowa- days [4], [6], [32] with cloud storage providers such as EMC, IBM, and Microsoft, offering products for multi- cloud systems [15], [16], [29].

In particular, we consider a system of s storage servers $S_1, \ldots, S_s$, and a collection of users. We assume that each server appropriately authenticates users. For simplicity and without loss of generality, we focus on the read/write storage abstraction of [21] which exports two operations:

- write(v) This routine splits v into s pieces $\{v_1, \ldots, v_s\}$ and sends (v_j) to server S_j, for $j \in [1 \ldots s]$.
- read() The read routine fetches the stored value v from the servers. For each $j \in [1 \ldots s]$, piece v_j is downloaded from server S_j and all

ISBN 978-93-88808-61-3

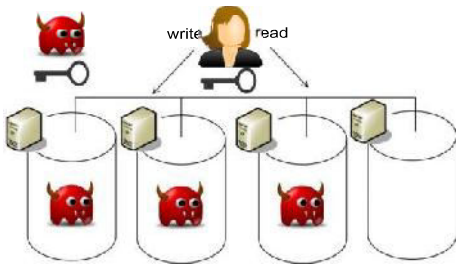


Fig. 1. Our attacker model. We assume an adversary which can acquire all the cryptographic secret material, and can compromise a large fraction (up to all but one) of the storage servers.

pieces are combined into v. We assume that the initial value of the storage is a special value $\perp$, which is not a valid input value for a write operation.

Adversarial Model

We assume a computationally-bounded adversary A which *can acquire the long-term cryptographic keys used to encrypt the data.* The adversary may do so either *(i)* by leveraging flaws or backdoors in the key-generation software [31], or *(ii)* by compromising the device that stores the keys (in the cloud or at the user). Since ciphertext blocks are distributed across servers hosted within different domains, we assume that the adversary cannot compromise all storage servers (cf. Figure 1).

In particular, we assume that the adversary can com- promise all but one of the servers and we model this adversary by giving it access to all but λ ciphertext blocks.

Note that if the adversary also learns the user's cre- dentials to log into the storage servers and downloads all the ciphertext blocks, then no cryptographic mech- anism can preserve data confidentiality. We stress that compromising the encryption key does not necessarily imply the compromise of the user's credentials. For example, encryption can occur on a specific-purpose device [10], and the key can be leaked, e.g., by the manufacturer; in this scenario, the user's credentials to access the cloud servers are clearly not compromised.

$(n-\lambda)$-CAKE Security

Existing security notions for encryption modes capture data confidentiality against an adversary which does not have the encryption key. That is, if the key is leaked, the confidentiality of data is broken.

In this paper we study an adversary that has access to the encryption key but does not have the entire ci- phertext. We therefore propose a new security definition that models our scenario.

As introduced above, we allow the adversary to access an encryption/decryption oracle and to "see" all but λ ciphertext blocks. Since confidentiality with $\lambda = 0$ isclearly not achievable[1], weinstead seekanencryption mode where $\lambda = 1$. However, having the flexibility of setting $\lambda \geq 1$ allows the design of more efficient schemes while keeping a high degree of security in practical deployments. (See Remark 7.)

We call our security notion $(n-\lambda)$ Ciphertext Access under Key Exposure, or $(n-\lambda)CAKE$. Similar to [12], $(n-\lambda)CAKE$ specifies a block length l such that a ciphertext y can be written as $y = y[1] \ldots y[n]$ where $|y[i]| = l$ and $n > 1$.

$$\begin{aligned}&\mathbf{Exp}^{(n-\lambda)CAKE}(A, b)\\ &\quad a \leftarrow \mathsf{K}(1^k)\\ &\quad x_0\, x_1\, state \leftarrow A^{\mathsf{E}^{F_a, F_a^{-1}}}(find)\\ &\quad y_b \leftarrow \mathsf{E}^{F_a F_a^{-1}}(x_b)\\ &\quad b' \leftarrow A^{Y_b, \mathsf{E}^{F_a, F_a^{-1}}}(guess, state)\end{aligned}$$

The adversary has unrestricted access to $\mathsf{E}^{F_a, F_a^{-1}}$ in both the "find" and "guess" stages. On input j, the oracle Y_b returns $y_b[j]$ and accepts up to $n-\lambda$ queries. On the one hand, unrestricted oracle access to $\mathsf{E}^{F_a, F_a^{-1}}$ captures the adversary's knowledge of the secret key. On the other hand, the oracle Y_b models the fact that the adversary has access to all but λ ciphertext blocks. This is the case when, for example, each server stores λ ciphertext blocks and the adversary cannot compromise all servers. The advantage of the adversary is defined as:

$$Adv^{(n-\lambda)CAKE}_{\mathsf{Q}}(A) = Pr[\mathbf{Exp}^{(n-\lambda)CAKE}_{\mathsf{Q}}(A, 1) = 1] - Pr[\mathbf{Exp}^{(n-\lambda)CAKE}_{\mathsf{Q}}(A, 0) = 1]$$

Definition 3. An encryption mode $= (\mathsf{K}, \mathsf{E}, \mathsf{D})$ is $(n-\lambda)CAKE$ secure if for any p.p.t. adversary A, we have $\mathrm{Adv}^{(n-\lambda)CAKE}_{\mathsf{Q}}(\mathrm{A}) \leq \varrho$, where ϱ is a negligible function in the security parameter.

Definition 3 resembles Definition 2 but has two fundamental differences. First, $(n-\lambda)CAKE$ refers to a keyed scheme and gives the adversary unrestricted access to the encryption/decryption oracles. Second, $(n-\lambda)CAKE$ relaxes the notion of all-or-nothing and parameterizes the number of ciphertext blocks that are not given to the adversary. As we will show in Sec- tion 4.2, this relaxation allows us to design encryption modes that are considerably more efficient than existing modes which offer a comparable level of security.

We stress that $(n-\lambda)CAKE$ does not consider confi- dentiality against "traditional" adversaries (i.e., adver- saries which do not know the encryption key). Indeed, an *ind*-adversary is not given the encryption key but has access to all ciphertext blocks. That is, the *ind*- adversary can compromise all the s storage servers. An $(n-\lambda)CAKE$-adversary is given the encryption key but can access all but λ ciphertext blocks. In practice,

1. Any party with access to all the ciphertext blocks and the encryption key can recover the plaintext.

ISBN 978-93-88808-61-3

the $(n - \lambda)$*CAKE*-adversary has the encryption key but can compromise up to $s - 1$ storage servers. Therefore, we seek an encryption mode with the following properties:

1) must be *ind* secure against an adversary which does not know the encryption key but has access to all ciphertext blocks (cf. Definition 1), by compro- mising all storage servers.
2) must be $(n - \lambda)$*CAKE* secure against an ad- versary which knows the encryption key but has access to $n - \lambda$ ciphertext blocks (cf. Definition 3), since it cannot compromise all storage servers.

REMARK 4. Property 2 ensures data confidentiality against the attacker model outlined in Section 3.2. Nevertheless, we must also account for weaker ad- versaries (i.e., traditional adversaries) that do not know the encryption key but can access the entire ciphertext —hence, *ind* security. Note that if the adversary which has access to the encryption key, can also access all the ciphertext blocks, then no cryptographic mechanism can preserve data confi- dentiality.

4 BASTION: SECURITY AGAINST KEY EXPOSURE

In this section, we present our scheme, dubbed Bastion, which ensures that plaintext data cannot be recovered as long as the adversary has access to all but *two* ciphertext blocks—even when the encryption key is exposed. We then analyze the security of Bastion with respect to Definition 1 and Definition 3.

Overview

Bastion departs from existing AON encryption schemes. Current schemes require a pre-processing round of block cipher encryption for the AONT, fol- lowed by another round of block cipher encryption (cf. Figure 2 (a)). Differently, Bastion first encrypts the data with one round of block cipher encryption, and then applies an efficient linear post-processing to the ciphertext (cf. Figure 2 (b)). By doing so, Bastion relaxes the notion of all-or-nothing encryption at the benefit of increased performance (see Figure 2).

More specifically, the first round of Bastion consists of CTR mode encryption with a randomly chosen key K, i.e., $y' = Enc(K, x)$. The output ciphertext y' is then fed to a linear transform which is inspired by the scheme of [28]. Namely, our transform basically computes $y = y' \cdot A$ where A is a square matrix such that: *(i)* all diagonal elements are set to 0, and *(ii)* the remaining off-diagonal elements are set to 1. As we shown later, such a matrix is invertible and has the nice property that $A^{-1} = A$. Moreover, $y = y' \cdot A$ ensures that each input block y'_j will depend on all output blocks y_i except from y_j. This transformation—combined with the fact that the original input blocks have high entropy (due to semantic secure encryption)—result in an *ind*-secure and $(n - 2)$*CAKE* secure encryption mode. In the following section, we show how to efficiently compute $y' \cdot A$ by means of bitwise XOR operations.

Bastion: Protocol Specification

We now detail the specification of Bastion.

On input a security parameter k, the key generation algorithm of Bastion outputs a key $K \in \{0, 1\}^k$ for the underlying block-cipher. Bastion leverages block cipher encryption in the CTR mode, which on input a plaintext bitstream x, divides it in blocks $x[1], \ldots, x[m]$, where m is odd[2] such that each block has size l.[3] The set of input blocks is encrypted under key K, resulting in ciphertext $y' = y'[1], \ldots, y'[m + 1]$, where $y'[m + 1]$ is an initialization vector which is randomly chosen from $\{0, 1\}^l$.

Next, Bastion applies a linear transform to y' as follows. Let $n = m + 1$ and assume A to be an n- by- n matrix where element $a_{i,j} = 0^l$ if $i = j$ or $a_{i,j} = 1^l$, otherwise.[4] Bastion computes $y = y' \cdot A$, where additions and multiplications are implemented by means of XOR and AND operations, respectively. That is, $y[i] \in y$ is computed as $y[i] = \bigoplus_{j=1}^{j=n} (y[j] \wedge a_{j,i})$, for $i = 1 \ldots, n$.

Given key K, inverting Bastion entails computing $y' = y \cdot A^{-1}$ and decrypting y' using K. Notice that matrix A is invertible and $A = A^{-1}$. The pseudocode of the encryption and decryption algorithms of Bastion are shown in Algorithms 1 and 2, respectively. Both algorithms use F to denote a generic block cipher (e.g., AES).

In our implementation, we efficiently compute the linear transform using $2n$ XOR operations as follows:

$$t = y'[1] \oplus y'[2] \oplus \cdots \oplus y'[n],\ y[i] = t \oplus y'[i],\ 1 \le i \le n.$$

Note that $y'[1] \ldots y'[n]$ (computed up to line 6 in Algorithm 1) are the outputs of the CTR encryption mode, where $y'[n]$ is the initialization vector. Similar to the CTR encryption mode, the final output of Bastion is one block larger than the original input.

Correctness Analysis

We show that for every $x \in \{0, 1\}^{lm}$ where m is odd, and for every $K \in \{0, 1\}^l$, we have $x = Dec(K, Enc(K,x))$.

In particular, notice that lines 2-6 of Algorithm 1 and lines 9-12 of Algorithm 2 correspond to the standard CTR encryption and decryption routines, respectively.

1. This requirement is essential for the correctness of the sub- sequent linear transform on the ciphertext blocks. That is, if m is even, then the transform is not invertible.
2. l is the block size of the particular block cipher used.
3. 0^l and 1^l denote a bitstring of l zeros and a bitstream of l ones, respectively.

ISBN 978-93-88808-61-3

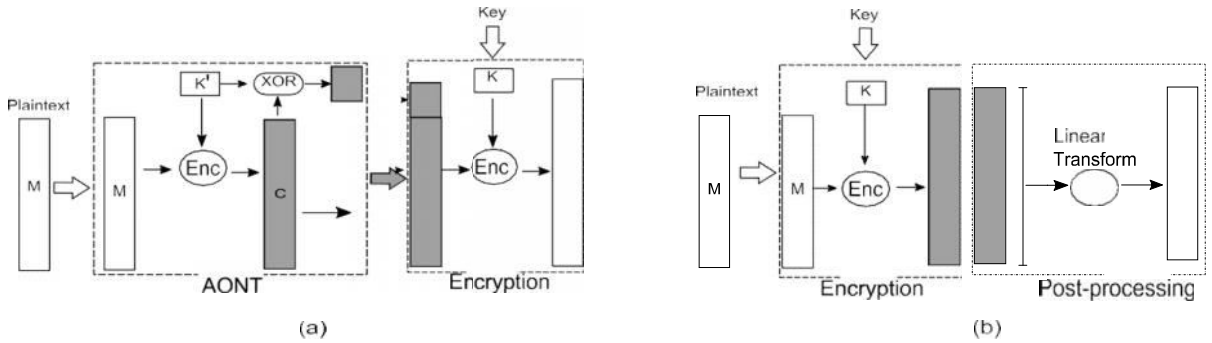


Fig. 2. (a) Current AON encryption schemes require a pre-processing round of block cipher encryption for the AONT, followed by another round of block cipher encryption. (b) On the other hand, Bastion first encrypts the data with one round of block cipher encryption, and then applies an efficient linear post-processing to the ciphertext.

Algorithm 1 Encryption in Bastion.

```
1: procedure Enc(K,x = x[1]...x[m])
2:     n = m + 1
3:     y'[n] ← {0, 1}^l                          ◁ y'[n] is the IV for CTR
4:     for i = 1 ... n − 1 do
5:         y'[i] = x[i] ⊕ F_K(y'[n]+i)
6:     end for
7:     t = 0^l
8:     for i = 1 ... n do
9:         t = t ⊕ y'[i]
10:    end for
11:    for i = 1 ... n do
12:        y[i] = y'[i] ⊕ t
13:    end for
14:    return y                                  ◁ y = y[1] ... y[n]
15: end procedure
```

Algorithm 2 Decryption in Bastion.

```
1:  procedure Dec(K, y = y[1] ... y[n])
2:  t = 0^l
3:  for i = 1 ... n do
4:      t = t ⊕ y[i]
5:  end for
6:  for i = 1 ... n do
7:      y'[i] = y[i]
⊕t 8: end for
9:  for i = 1 ... n − 1 do
10:     x[i] = y'[i] ⊕ F_K(y'[n] + i)
11: end for
12:     return x                                  ◁ x = x[1] ... x[n − 1]
13: end procedure
```

Therefore, we are only left to show that the linear transformation computed in lines 7-14 of Algorithm 1 is correctly reverted in lines 2-8 of Algorithm 2. In other words, we need to show that $t = \bigoplus_{i=1..n} y[i]$ (as computed in the decryption algorithm) matches $t = \bigoplus_{i=1..n} y[i]'$ (as computed in the encryption algo- rithm).

Recall that t can be computed as follows:

$$t = \bigoplus_{i=1..n} y[i] = \bigoplus_{i=1..n} (y'[i] \oplus t) = \bigoplus_{i=1..n} y'[i] \oplus \bigoplus_{i=1..n} y'[i] \oplus \bigoplus_{i=1..n} \bigoplus_{j=1..n, j \neq i} y'[j] = \bigoplus_{i=1..n} y'[i]$$

Notice that the last step holds because n is even and therefore each $y'[j]$ is XORed for an odd number of times.

REMARK 5. We point out that Bastion is not restricted to the CTR encryption mode and can be instanti- ated with other ind-secure block cipher (and stream ciphers) modes of encryption (e.g., CBC, OFB).

Tointerface with our cloud storage model described in Section 3.1, we assume that each user encrypts the data using Bastion before invoking the write() routine. More specifically, let $\mathrm{Enc}(K, \cdot)$, $\mathrm{Dec}(K, \cdot)$ denote the encryption and decryption routines of Bastion, respectively. Given encryption key K and a file f, the user computes $v \leftarrow \mathrm{Enc}(K, f)$ and invokes write(v) in order to upload the encrypted file to the cloud. In this setting, key K remains stored at the user'smachine. Similarly, to download the file from the cloud, the user invokes read($\cdot$) to fetch v and runs $f \leftarrow \mathrm{Dec}(K, v)$ to recover f.

Security Analysis

In this section, we show that Bastion is *mathrmind* secure and $(n - 2)$*CAKE* secure.

LEMMA 1. Bastion is ind secure.

Proof 1. Bastion uses an *ind* secure encryption mode to encrypt a message, and then applies a linear

ISBN 978-93-88808-61-3

transform on the ciphertext blocks. It is straight- forward to conclude that Bastion is *ind* secure. In other words, a polynomial-time algorithm A that has non-negligible advantage in breaking the *ind* security of Bastion can be used as a black-box by another polynomial-time algorithm B to break the *ind* security of the underlying encryption mode. In particular, B forwards A's queries to its oracle and applies the linear transformation of Algorithm 1 lines 7-14 to the received ciphertext before forward- ing it to A. The same strategy is used when A outputs two messages at the end of the *find* stage: the two messages are forwarded to B's oracle; upon receiving the challenge ciphertext, B applies the linear transformation and forwards it to A. When A replies with its guess b', B outputs the same guess. It is easy to see that if A has non-negligible advantage in guessing correctly which message was encrypted, so does B. Furthermore, the running time of B is the one of A plus the time to apply the linear transformation to A's queries.

LEMMA 2. Given any $n - 2$ blocks of $y[1] \ldots y[n]$ as output by Bastion, it is infeasible to compute any $y'[i]$, for $1 \leq i \leq n$.

Proof 2. Let $y = y[1], \ldots, y[n] \leftarrow E(K, x = x[1] \ldots x[m])$. Note that given any $(n - 1)$ blocks of y, the adversary can compute one block of y'. In particular, $y'[i] = \bigoplus_{j=1, j \neq i}^{n} y[j]$, for any $1 \leq i \leq n$. As it will become clear later, with one block $y'[i]$ and the encryption key, the adversary has non-negligible probability of winning the game of Definition 3. However, if only $(n - 2)$ blocks of y are given, then each of the n blocks of y' can take on any possible values in $\{0, 1\}^l$, depending on the two unknown blocks of y. Recall that each block $y'[i]$ is dependent on $(n - 1)$ blocks of y and it is pseudo-random as output by the CTR encryption mode. Therefore, given any $(n - 2)$ blocks of y, then $y'[i]$ could take any of the 2^l possibilities, for $1 \leq i \leq n$.

LEMMA 3. Bastion is $(n - 2)$*CAKE* secure.

Proof 3. The security proof of Bastion resembles the standard security proof of the CTR encryption mode and relies on the existence of pseudo-random permutations. In particular, given a polynomial-type algorithm A which has non-negligible advantage in the $(n - \lambda)$*CAKE* experiment with $\lambda = 2$, we can construct a polynomial-time algorithm B which has non-negligible advantage in distinguishing between a true random permutation and a pseudo-random permutation.

B has access to oracle O and uses it to answer the encryption and decryption queries issued by A. In particular, A's queries are answered as follows:

- *Decryption query for* $y[1] \ldots y[n]$
  1) Compute $t = y[1] \oplus \ldots \oplus y[n]$
  2) Compute $y'[i] = y[i] \oplus t$, for $1 \leq i \leq n$
  3) Compute $x[i] = y'[i] \oplus O(y'[n] + i)$, for $1 \leq i \leq n - 1$
  4) Return $x[1] \ldots x[n - 1]$
- *Encryption query for* $x[1] \ldots x[n - 1]$
  1) Pick random $y'[n] \in \{0, 1\}^l$
  2) Compute $y'[i] = x[i] \oplus O(y'[n] + i)$, for $1 \leq i \leq n - 1$
  3) Compute $t = y'[1] \oplus \ldots \oplus y'[n]$
  4) Compute $y[i] = y'[i] \oplus t$, for $1 \leq i \leq n$
  5) Return $y[1] \ldots y[n]$

When A outputs two messages $x_1[1] \ldots x_1[n-1]$ and $x_2[1] \ldots x_2[n - 1]$, B picks $b \in \{0, 1\}$ at random and does the following:

1) Pick random $y'_b[n] \in \{0, 1\}^l$
2) Compute $y'_b[i] = x_b[i] \oplus O(y'_b[n], i)$, for $1 \leq i \leq n-1$
3) Compute $t = y_b[1] \oplus \ldots \oplus y_b[n]$
4) Compute $y_b[i] = y'_b[i] \oplus t$, for $1 \leq i \leq n$

At this point, A selects $(n - 2)$ indexes $i_1, \ldots i_{n-2}$ and B returns the corresponding $y_b[i_1], \ldots, y_b[i_{n-2}]$. Encryption and decryption queries are answered as above. When A outputs its answer b', B outputs 1 if $b = b'$, and 0 otherwise. It is straightforward to see that if A has advantage larger than negligible to guess b, then B has advantage larger than negligible to distinguish a true random permutation from a pseudorandom one. Furthermore, the number of queries issued by B to its oracle amounts to the number of encryption and decryption queriesissued by A. Note that by Lemma 2, during the guess stage, A cannot issue a decryption query on the challenge ciphertext since with only $(n - 2)$ blocks, finding the remaining blocks is infeasible.

REMARK 6. Bastion is not $(n - 1)$*CAKE* secure. As shown in the proof of Lemma 2, the adversary can recover one block of y' given any $(n - 1)$ blocks of y. If the adversary recovers $y'[n]$ that is used as an IV in the CTR encryption mode, the adversary can easily win the $(n - 1)$*CAKE* game. Recall that our security definition allows the adversary to learn the encryption key.

REMARK 7. Bastion is $(n - 2)$*CAKE* secure according to Definition 3. However, in a practical deployment, we expect that each file spans several thousands blocks [5]. When those blocks are evenly spread across servers, each server will store a larger number of blocks. Therefore, an $(n - 2)$*CAKE* secure scheme such as Bastion clearly preserves data confidential- ity unless *all* servers are compromised.

4. For example, a 10MB file encrypted using AES has more than 600K blocks.

TABLE 1
Comparison between Bastion and existing constructs. We assume a plaintext of m = n − 1 blocks. Since all schemes are symmetric, we only show the computation overhead for the encryption/encoding routine in the column "Computation" ("b.c." is the number of block cipher operations; "XOR" is the number of XOR operations.

	Computation	**Storage** (blocks)	**Security**
CTR Encryption	n − 1 b.c. n − 1 XOR	n	1*CAKE* ind-secure
Rivest AONT [26]	2(n − 1) b.c. 3(n − 1) XOR	n	N/A ind-INsecure
Desai AONT [12]	n − 1 b.c. 2(n − 1) XOR	n	N/A ind-INsecure
Rivest AON Encryption [26]	3n − 2 b.c. 3(n − 1) XOR	n	(n − 1)*CAKE* ind-secure
Desai AON Encryption [12]	2n − 1 b.c. 2(n − 1) XOR	n	(n − 1)*CAKE* ind-secure
Encrypt-then -secret-share	n − 1 b.c. 2n − 1 XOR	n^2	(n−1)*CAKE* ind-INsecure*
Bastion	n − 1 b.c. 3n − 1 XOR	n	(n − 2)*CAKE* ind-secure

* Recall that an *ind*-adversary can access all storage servers to fetch all ciphertext blocks. Therefore, the adversary can also fetch all the key shares and compute the encryption key.

5 Comparison to Existing Schemes

In what follows, we briefly overview several encryption modes and argue about their security (according to Definitions 1 and 3) and performance when compared to Bastion.

CPA-encryption modes

Traditional CPA-encryption modes, such as the CTR mode, provide *ind* security but are only 1*CAKE* secure. That is, an adversary equipped with the encryption key must only fetch two ciphertext blocks to break data confidentiality.[6]

CPA-encryption and secret-sharing

Another option is to rely on the combination of CPA secure encryption modes and secret-sharing.

If the file f is encrypted and then shared with an n-out-of-n secret-sharing scheme(denoted as "encrypt- then-secret-share" in the following), then the construc- tion is clearly $(n-1)$*CAKE* secure and is also *ind* secure. However, secret-sharing the ciphertext comes at considerable storage costs; for example, each share would be as large as the file f using a perfect secret sharing scheme—which makes it impractical for storing large files.

Secret-sharing the encryption key and dispersing its shares across the storage servers alongside the cipher- text is not secure against an *ind*-adversary. Indeed, if the adversary can access all the storage servers and down- load all ciphertext blocks, the adversary may as well download all key shares and compute the encryption key.

1. We assume that the CTR encryption routine starts with a random IV that is incremented at every block encryption.

AON encryption

Recall that an AONT is not an encryption scheme and does not require the decryptor to have any secret key. That is, an AONT is not secure against an *ind*-adversary which can access all the ciphertext blocks. One alter- native is to combine the use of AONT with standard encryption. Rivest [26] suggests to pre-process a mes- sage with an AONT and then encrypt its output with an encryption mode. This paradigm is referred to in the literature as AON encryption and provides $(n-1)$*CAKE* security. Existing AON encryption schemes require at least two rounds of block cipher encryption with two different keys [12], [26]. At least one round is required for the actual AONT that embeds the first encryption key in the pseudo-ciphertext (cf. Section 2). An addi- tional round uses another encryption key that is kept secret to guarantee CPA-security. However, two encryp- tion rounds constitute a considerable overhead when encrypting and decrypting large files. In Appendix A, we describe possible ways of modifying the AONTs of [26] and [12] to achieve *ind* security and $(n-1)$*CAKE* security without adding another round of block cipher encryption, and we discuss their shortcomings.

Clearly, these solutions are either not satisfactory in terms of security or incur a large overhead when compared to Bastion and may not be suitable to store large files in a multi-cloud storage system.

Performance Comparison

Table 1 compares the performance of Bastion with the encryption schemes considered so far, in terms of computation, storage, and security.

Given a plaintext of m blocks, the CTR encryption mode outputs $n = m + 1$ ciphertext blocks, computed with $(n-1)$ block cipher operations and $(n-1)$ XOR

operations. The CTR encryption mode is *ind* secure but only 1*CAKE* secure.

Rivest AONT outputs a pseudo-ciphertext of $n = m + 1$ blocks using $2(n - 1)$ block cipher operations and $3(n-1)$ XOR operations. Desai AONT outputs the same number of blocks but requires only $(n - 1)$ block cipher operations and $2(n - 1)$ XOR operations. Both Rivest AONT and Desai AONT are, however, not *ind* secure since the encryption key used to compute the AONT output is embedded in the output itself. Encrypting the output of Rivest AONT or Desai AONT with a stan- dard encryption mode (both [12] and [26] use the ECB encryption mode), requires additional n block cipher operations, and yields an AON encryption that is *ind* secure[7] and $(n - 1)$*CAKE* secure. Encrypt-then-secret- share (cf. Section 4.4) is *ind* secure and $(n - 1)$*CAKE* secure. It requires $(n - 1)$ block cipher operations and n XOR operations if additive secret sharing is used. How- ever secret-sharing encryption results in a prohibitively large storage overhead of n^2 blocks.

Bastion also outputs $n = m + 1$ ciphertext blocks. It achieves *ind* security and $(n - 2)$*CAKE* security with only $(n - 1)$ block cipher operations and $(3n - 1)$ XOR operations.[8]

We conclude that Bastion achieves a solid tradeoff between the computational overhead of existing AON encryption modes and the exponential storage overhead of secret-sharing techniques, while offering a compa- rable level of security. In Section 6, we confirm the superior performance of Bastion by means of imple- mentation.

6 Implementation and Evaluation

In this section, we describe and evaluate a prototype implementation modeling a read-write storage system based on Bastion. We also discuss insights with respect to the integration of Bastion within existing dispersed storage systems.

Implementation Setup

Our prototype, implemented in C++, emulates the read- write storage model of Section 3.1. We instantiate Bastion with the CTR encryption mode (cf. Figure 1) using both AES128 and Rijndael256, implemented using the libmcrypt.so. 4.4.7 library. Since this library doesnot natively support the CTR encryption mode, we use it for the generation of the CTR keystream, which is later XORed with the plaintext.

We compare Bastion with the AON encryption schemes of Rivest [26] and Desai [12]. For baseline comparison, we include in our evaluation the CTR encryption mode and the AONTs due to Rivest [26] and Desai [12], which are used in existing dispersed storage systems, e.g., Cleversafe [25]. We do not evaluate the performance of secret-sharing the data because of its prohibitively large storage overhead (squared in the number of input blocks). We evaluate our implemen- tations on an Intel(R) Xeon(R) CPU E5-2470 running at 2.30GHz. Note that the processor clock frequency might have been higher during the evaluation due to the TurboBoost technology of the CPU. In our evaluation, we abstract away the effects of network delays and congestion, and we only assess the processing perfor- mance of the encryption for the considered schemes. This is a reasonable assumption since all schemes are length-preserving (plus an additional block of l bits), and are therefore likely to exhibit the same network performance. Moreover, we only measure the per- formance incurred during encryption/encoding, since all schemes are symmetric, and therefore the decryp- tion/decoding performance is comparable to that of the encryption/encoding process.

We measure the peak throughput and the latency ex- hibited by our implementations w.r.t. various file/block sizes. For each data point, we report the average of 30 runs. Due to their small widths, we do not show the corresponding 95% confidence intervals.

Evaluation Results

Our evaluation results are reported in Figure 3 and Figure 4. Both figures show that Bastion considerably improves (by more than 50%) the performance of ex- isting $(n - 1)$*CAKE* encryption schemes and only in- curs a negligible overhead when compared to existing semantically secure encryption modes (e.g., the CTR encryption mode) that are only 1*CAKE* secure.

In Figure 3, we show the peak throughput achie- ved by the CTR encryption mode, Bastion, Desai AONT/AON, and Rivest AONT/AON schemes. The peak throughput achieved by Bastion reaches almost 72 MB/s and is only 1% lower than the one exhibited by the CTR encryption mode. When compared with ex- isting $(n - 1)$*CAKE* secure schemes, such as Desai AON encryption and Rivest AON encryption, our results show that the peak throughput of Bastion is almost twice as large as that of Desai AON encryption, and more than three times larger than the peak throughput of Rivest AON encryption.

We also evaluate the performance of Bastion, with respect to different block sizes of the underlying block cipher. Our results show that—irrespective of the block size—Bastion only incurs a negligible performance de- terioration in peak throughput when compared to the CTR encryption mode. Figures 4(a) and 4(b) show the latency (in ms) incurred by the encryption/encoding routines for different file sizes. The latency of Bastion is comparable to that of the CTR encryption mode—for both AES128 and Rijandael256—and results in a con- siderable improvement over existing AON encryption schemes (more than 50% gain in latency).

1. Security according to Definition 1 is achieved because the key used to create the AONT is always random, even if the key used to add the outer layer of encryption is fixed.

2. Bastion requires (n− 1) XOR operations for the CTR encryp- tion and 2n XOR operations for the linear transform.

ISBN 978-93-88808-61-3

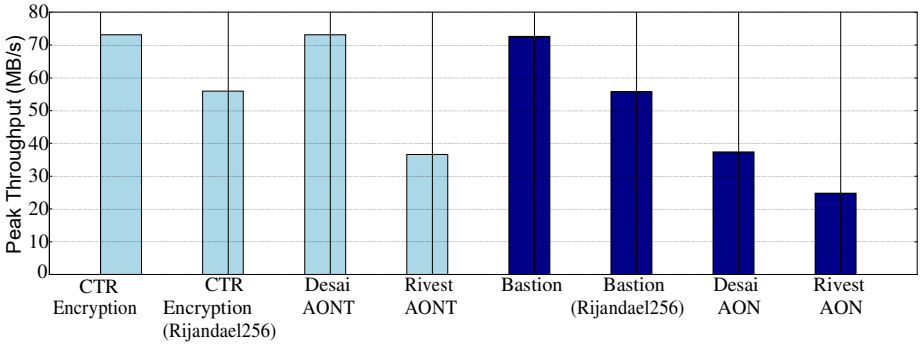
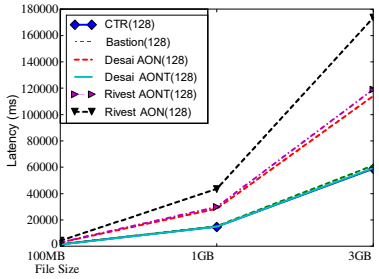
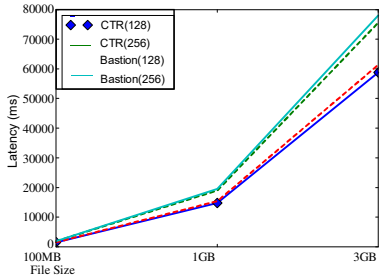


Fig. 3. Peak throughput comparison. Unless otherwise specified, the underlying block cipher is AES128. Each data point is averaged over 30 runs. Histograms in dark blue depict encryption modes which offer comparable security to Bastion. Light blue histograms refer to encryption/encoding modes where individual ciphertext blocks can be inverted when the key is exposed.



(a) Latency of encryption/encoding for differ- ent file sizes.



(b) Latency of encryption/encoding for different block sizes of the underlying block cipher.

Fig. 4. Performance evaluation of Bastion. Each data point in is averaged over 30 runs. Unless otherwisespecified, the underlying block cipher is AES-128. CTR(256) and Bastion(256) denote the CTR encryption mode and Bastion encryption routine, respectively, instantiated with Rijandael256.

Deployment within HYDRAstor

Recall that Bastion preserves data confidentiality against an adversary that has the encryption key as long as the adversary does not have access to two ciphertext blocks. In a multi-cloud storage system, if each server stores at least two ciphertext blocks, then Bastion clearly preserves data confidentiality unless *all* servers are compromised.

In scenarios where servers can be faulty, Bastion can be combined with information dispersal algorithms (e.g., [24]) to provide data confidentiality and fault tolerance. Recall that information dispersal algorithms (IDA), parameterized with t_1, t_2 (where $t_1 \leq t_2$), encode data into t_2 symbols such that the original data can be recovered from any t_1 encoded symbols. In our multi- cloud storage system (cf. Section 3.1), the ciphertext output by Bastion is then fed to the IDA encoding routine, with symbols of size l bits, and with parameters $t_2 \geq 2s$, $t_1 < t_2$, where s is the number of available servers. Since the output of the IDA is equally spread across the s servers, by setting $t_2 \geq 2s$, we ensure that each server stores at least two ciphertext blocks worth of data. Finally, the encoded symbols are input to the write() routine that distributes symbols evenly to each of the storage servers. Recovering f via the read() routine entails fetching t_1 encoded symbols from the servers and decoding them via the IDA decoding routine. The resulting ciphertext can be decrypted using Bastion to recover file f. By doing so, data confiden- tiality is preserved even if the key is exposed unless

$t = \frac{st_1}{t_2}$ servers are compromised. Furthermore, data availability is guaranteed in spite of $(s - t)$ server failures.

HYDRAstor

We now discuss the integration of a prototype implementation of Bastion within the HYDRAstor grid storage system [13], [23]. HYDRAstor is a commer- cial secondary storage solution for enterprises, which consists of a back-end architectured as a grid of stor- age nodes built around a distributed hash table. HY- DRAstor tolerates multiple disk, node and network failures, rebuilds the data automatically after failures, and informs users about recoverability of the deposited data [13]. The reliability and availability of the stored data can be dynamically adjusted by the clients with each write operation, as the back-end supports multiple data resiliency classes [13].

HYDRAstor distributes written data to multiple disks using the distributed resilient data technology (DRD); the combination of Bastion with DRD ensures that an adversary which has the encryption key and compromises a subset of the disks (i.e., determined by the reconstruction threshold), cannot acquire any mean- ingful information about the data stored on the disk. To better assess the performance impact of Bastion in HYDRAstor, we evaluated the performance of Bastion in the newest generation HYDRAstor HS8-4000 series system, which uses CPUs with accelerated AES encryption (i.e., the AESNI instruction set). In our experiments, all written data was unique to remove the effect of data deduplication. Results show that the write bandwidth was not affected by the integration of Bastion. The read bandwidth decreased only by 3%. In both read and write operations, the CPU utilization in the system only increased marginally. These experiments clearly suggest that Bastion can be integrated in existing com- mercial storage systems to strengthen the security of these systems under key exposure, without affecting performance.

7 Related Work

To the best of our knowledge, this is the first work that addresses the problem of securing data stored in multi- cloud storage systems when the cryptographic material is exposed. In the following, we survey relevant related work in the areas of deniable encryption, information dispersal, all-or-nothing transformations, secret-sharing techniques, and leakage-resilient cryptography.

Deniable Encryption

Our work shares similarities with the notion of "shared- key deniable encryption" [9], [14], [18]. An encryption scheme is "deniable" if—when coerced to reveal the en- cryption key—the legitimate owner reveals "fake keys" thus forcing the ciphertext to "look like" the encryption of a plaintext different from the original one—hence keeping the original plaintext private. Deniable en- cryption therefore aims to deceive an adversary which does not know the "original" encryption key but, e.g., can only acquire "fake" keys. Our security definition models an adversary that has access to the real keying material.

Information Dispersal

Information dispersal based on erasure codes [30] has been proven as an effective tool to provide reliability in a number of cloud-based storage systems [1], [2], [20], [33]. Erasure codes enable users to distribute their data on a number of servers and recover it despite some servers failures.

Ramp schemes [7] constitute a trade-off between the security guarantees of secret sharing and the efficiency of information dispersal algorithms. A ramp scheme achieves higher "code rates" than secret sharing and features two thresholds t_1, t_2. At least t_2 shares are required to reconstruct the secret and less than t_1 shares provide no information about the secret; a number of shares between t_1 and t_2 leak "some" information.

All or Nothing Transformations

All-or-nothing transformations (AONTs) were first introduced in [26] and later studied in [8], [12]. The majority of AONTs leverage a secret key that is em- bedded in the output blocks. Once all output blocks are available, the key can be recovered and single blocks can be inverted. AONT, therefore, is not an encryption scheme and does not require the decryptor to have any key material. Resch et al. [25] combine AONT and information dispersal to provide both fault-tolerance and data secrecy, in the context of distributed storage systems. In [25], however, an adversary which knows the encryption key can decrypt data stored on single servers.

Secret Sharing

Secret sharing schemes [5] allow a dealer to distribute a secret among a number of shareholders, such that only authorized subsets of shareholders can reconstruct the secret. Inthresholdsecret sharing schemes[11], [27], the dealer defines a threshold t and each set of shareholders of cardinality equal to or greater than t is authorized to reconstruct the secret. Secret sharing guarantees se- curity against a non-authorized subset of shareholders; however, they incur a high computation/storage cost, which makes them impractical for sharing large files. Rabin [24] proposed an information dispersal algorithm with smaller overhead than the one of [27], however the proposal in [24] does not provide any security guarantees when a small number of shares (less than the reconstruction threshold) are available. Krawczyk
[19] proposed to combine both Shamir's [27] and Ra- bin's [24] approaches; in [19] a file is first encrypted using AES and then dispersed using the scheme in [24], while the encryption key is shared using the scheme in [27]. In Krawczyk's scheme, individual ciphertext blocks encrypted with AES can be decrypted once the key is exposed.

Leakage-resilient Cryptography

Leakage-resilient cryptography aims at designing cryptographic primitives that can resist an adversary which learns partial information about the secret state of a sys- tem, e.g., through side-channels [22]. Different models allow to reason about the "leaks" of real implemen- tations of cryptographic primitives [22]. All of these models, however, limit in some way the knowledge of the secret state of a system by the adversary. In contrast, the adversary is given all the secret material in our model.

8 CONCLUSION

In this paper, we addressed the problem of securing data outsourced to the cloud against an adversary which has access to the encryption key. For that pur- pose, we introduced a novel security definition that captures data confidentiality against the new adversary. We then proposed Bastion, a scheme which ensures the confidentiality of encrypted data even when the adversary has the encryption key, and all but *two* cipher- text blocks. Bastion is most suitable for settings where the ciphertext blocks are stored in multi-cloud storage systems. In these settings, the adversary would need to acquire the encryption key, and to compromise *all* servers, in order to recover any single block of plaintext.

We analyzed the security of Bastion and evaluated its performance in realistic settings. Bastion consider- ably improves (by more than 50%) the performance of existing primitives which offer comparable secu- rity under key exposure, and only incurs a negligible overhead (less than 5%) when compared to existing semantically secure encryption modes (e.g., the CTR encryption mode). Finally, we showed how Bastion can be practically integrated within existing dispersed storage systems.

REFERENCES

[1] M. Abd-El-Malek, G. R. Ganger, G. R. Goodson, M. K. Re- iter, and J. J. Wylie, "Fault-Scalable Byzantine Fault-Tolerant Services," in *ACM Symposium on Operating Systems Principles (SOSP)*, 2005, pp. 59–74.

[2] M. K. Aguilera, R. Janakiraman, and L. Xu, "Using Erasure Codes Efficiently for Storage in a Distributed System," in *International Conference on Dependable Systems and Networks (DSN)*, 2005, pp. 336–345.

[3] W. Aiello, M. Bellare, G. D. Crescenzo, and R. Venkatesan, "Security amplification by composition: The case of doubly-iterated, ideal ciphers," in *Advances in Cryptology (CRYPTO)*, 1998, pp. 390–407.

[4] C. Basescu, C. Cachin, I. Eyal, R. Haas, and M. Vukolic, "Ro- bust Data Sharing with Key-value Stores," in *ACM SIGACT- SIGOPS Symposium on Principles of Distributed Computing (PODC)*, 2011, pp. 221–222.

[5] A. Beimel, "Secret-sharing schemes: A survey," in *Interna- tional Workshop on Coding and Cryptology (IWCC)*, 2011, pp. 11–46.

[6] A. Bessani, M. Correia, B. Quaresma, F. André, and P. Sousa, "DepSky: Dependable and Secure Storage in a Cloud-of- clouds," in *Sixth Conference on Computer Systems (EuroSys)*, 2011, pp. 31–46.

[7] G. R. Blakley and C. Meadows, "Security of ramp schemes," in *Advances in Cryptology (CRYPTO)*, 1984, pp. 242–268.

[8] V. Boyko, "On the Security Properties of OAEP as an All- or-nothing Transform," in *Advances in Cryptology (CRYPTO)*, 1999, pp. 503–518.

[9] R. Canetti, C. Dwork, M. Naor, and R. Ostrovsky, "Deniable Encryption," in *Proceedings of CRYPTO*, 1997.

[10] Cavalry, "Encryption Engine Dongle," http://www. cavalrystorage.com/en2010.aspx/.

[11] C. Charnes, J. Pieprzyk, and R. Safavi-Naini, "Conditionally secure secret sharing schemes with disenrollment capability," in *ACM Conference on Computer and Communications Security (CCS)*, 1994, pp. 89–95.

[12] A. Desai, "The security of all-or-nothing encryption: Protect- ing against exhaustive key search," in *Advances in Cryptology (CRYPTO)*, 2000, pp. 359–375.

[13] C. Dubnicki, L. Gryz, L. Heldt, M. Kaczmarczyk, W. Kil- ian, P. Strzelczak, J. Szczepkowski, C. Ungureanu, and M. Welnicki, "HYDRAstor: a Scalable Secondary Storage," in *USENIX Conference on File and Storage Technologies (FAST)*, 2009, pp. 197–210.

[14] M. Dürmuth and D. M. Freeman, "Deniable encryption with negligible detection probability: An interactive construction," in *EUROCRYPT*, 2011, pp. 610–626.

[15] EMC, "Transform to a Hybrid Cloud," http://www.emc. com/campaign/global/hybridcloud/index.htm.

[16] IBM, "IBM Hybrid Cloud Solution," http://www-01.ibm. com/software/tivoli/products/hybrid-cloud/.

[17] J. Kilian and P. Rogaway, "How to protect DES against exhaustive key search," in *Advances in Cryptology (CRYPTO)*, 1996, pp. 252–267.

[18] M. Klonowski, P. Kubiak, and M. Kutylowski, "Practical De-niable Encryption," in *Theory and Practice of Computer Science (SOFSEM)*, 2008, pp. 599–609.

[19] H. Krawczyk, "Secret Sharing Made Short," in *Advances in Cryptology (CRYPTO)*, 1993, pp. 136–146.

[20] J. Kubiatowicz, D. Bindel, Y. Chen, S. E. Czerwinski, P. R. Eaton, D. Geels, R. Gummadi, S. C. Rhea, H. Weatherspoon, W. Weimer, C. Wells, and B. Y. Zhao, "OceanStore: An Archi-tecture for Global-Scale Persistent Storage," in *International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, 2000, pp. 190–201.

[21] L. Lamport, "On interprocess communication," 1985.

[22] S. Micali and L. Reyzin, "Physically observable cryptography (extended abstract)," in *Theory of Cryptography Conference (TCC)*, 2004, pp. 278–296.

[23] NEC Corp., "HYDRAstor Grid Storage," http://www. hydrastor.com.

[24] M. O. Rabin, "Efficient dispersal of information for security, load balancing, and fault tolerance," *J. ACM*, vol. 36, no. 2, pp. 335–348, 1989.

[25] J. K. Resch and J. S. Plank, "AONT-RS: Blending Security and Performance in Dispersed Storage Systems," in *USENIX Conference on File and Storage Technologies (FAST)*, 2011, pp. 191–202.

[26] R. L. Rivest, "All-or-Nothing Encryption and the Package Transform," in *International Workshop on Fast Software Encryp-tion (FSE)*, 1997, pp. 210–218.

[27] A. Shamir, "How to Share a Secret?" in *Communications of the ACM*, 1979, pp. 612–613.

[28] D. R. Stinson, "Something About All or Nothing (Trans- forms)," in *Designs, Codes and Cryptography*, 2001, pp. 133– 138.

[29] StorSimple, "Cloud Storage," http://www.storsimple.com/.

[30] J. H. van Lint, *Introduction to Coding Theory*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 1982.

[31] Wikipedia, "Edward Snowden," http://en.wikipedia.org/ wiki/Edward_Snowden#Disclosure.

[32] Z. Wu, M. Butkiewicz, D. Perkins, E. Katz-Bassett, and H. V. Madhyastha, "SPANStore: Cost-effective Geo-replicated Stor- age Spanning Multiple Cloud Services," in *ACM Symposium on Operating Systems Principles (SOSP)*, 2013, pp. 292–308.

[33] H. Xia and A. A. Chien, "RobuSTore: a Distributed Stor- age Architecture with Robust and High Performance," in *ACM/IEEE Conference on High Performance Networking and Computing (SC)*, 2007, p. 44.

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

Data Implementation on Friendly Environment with Green Data Mining Process

Tabitha indupalli
tabi.indupalli@gmail.com

Sharon Drisilda
emmanuelalfred14@gmail.com

Abstract

This paper develops a set of principles for green data mining, related to the key stages of business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The principles are grounded in a review of the Cross Industry Standard Process for Data mining (CRISP-DM) model and relevant literature on data mining methods and Green IT. We describe how data scientists can contribute to designing environmentally friendly data mining processes, for instance, by using green energy, choosing between make-or-buy, exploiting approaches to data reduction based on business understanding or pure statistics, or choosing energy friendly models.

1. Introduction

The use of computing power coupled with the unprecedented availability of data provide ample opportunity to improve energy efficiency . However, they are also an increasingly relevant source of energy consumption and associated carbon emissions. Data centers consumed about 70 billion kWh in 2016 in the United States alone , and the total consumption of all IT is estimated to be close to 5% of total energy consumption. In response to this increasing amount of energy used by IT, Greenpeace published the "Guide to Building the Green Internet" , promoting "a more widespread adaption in best practices" for energy efficient data center design. They demand that "data center operators and customers should regularly report their energy performance and establish transparent energy savings targets." Electricity consumption is costly—it involves various detrimental effects on nature and society, ranging from bird deaths by wind turbines, on to severe air pollution and CO2 emissions by coal power plants, and the risk of catastrophes stemming from nuclear power plants.

These concerns are partially addressed by current initiatives under notions such as green information systems (Green IS) or green information technology (Green IT) , but environmentally friendly data mining is a novel topic.

Data scientists often leverage a large pool of computational resources using sophisticated and computationally costly machine learning techniques to extract knowledge and insights from data. Though existing processes such as the Cross Industry Standard Process for Data mining (CRISP-DM) provide some guidance on how to execute a data mining project, the skills of a data scientist heavily rely on creativity involving many degrees of freedom, often including the choice of tools, models, and data sources.

It is against this background that, in this paper, we develop guidelines for data scientists to implement more environmentally friendly practices that can complement technology-focused perspectives aiming to design more energy efficient IT-based systems. Specifically, we are focusing attention on one important area of data science—data mining. Data mining can be described as knowledge discovery from data or in terms of different activities as collecting, cleaning, processing, analyzing and gaining useful insights from data . We ask: How can data *scientists implement more environmentally friendly data mining processes?*

The remainder of this paper is structured as follows. We first describe our methodology. We then review the data mining process and develop a set of principles for green data mining. We conclude by discussing limitations and future work.

2. Methodology

We derived our principles by analyzing the CRISP-DM data mining process and literature on green IT and data mining. In a first step, we identified factors determining energy consumption. In a second step, we identified individual steps of the CRISP-DM process by investigating possibilities for reduction of each factor. We limited our analysis to those aspects that can be directly influenced by data scientists, including the choice of data, its representation, as well as processes and techniques used throughout the data

analysis process. We do not target the development of novel data mining algorithms for specific problems or improving hardware or software, though some of our insights might be helpful in guiding such developments.

We conducted a narrative literature review on green IT, green IS, and data mining because our goal was to investigate elementary factors and research outcomes related to these areas of research. Green data science is a novel field and, therefore, is more amenable to a qualitative approach such as narrative literature review than a more quantitative approach detailing the current-state-of-research, as done for a descriptive review. Our focus was on using established online databases from computer science as well as information systems such as IEEE Xplore, ProQuest (ABI/INForm), ScienceDirect (Elsevier), AIS electronic library and the ACM digital library. We did not limit ourselves to journals since new ideas are often presented first at academic conferences and a significant body of works, in particular in the field of computer science, only appear as conference articles.

3. The data mining process

There are multiple data mining processes , most of which share common phases. CRISP-DM is arguably the most widely known and practiced model , attending to business and data understanding, data preparation, modelling, evaluation and deployment (Figure 1). The business understanding phase clarifies project objectives and business requirements, which are then translated into a data mining problem. There are unsupervised data mining problems including association pattern mining and clustering as well as supervised approaches like classification . Data understanding typically requires initial data selection or collection. Data is first analyzed in an exploratory fashion to get a basic understanding of the data in the business context. Exploratory analysis supports the development of hypothesis by identifying patterns in the data [3]. It allows to get first insights as well as to identify data quality problems. Data preparation includes using raw data to derive data that can be fed into the models. Activities include data selection, transformation, and cleaning. The data might have to be prepared separately for each model. The modelling phase consists of defining suitable models, selecting a model, and adapting the model, for instance, optimizing its parameters to solve the data mining problem. Computational evaluation of the model is part of the model selection process. Every data mining problem can be tackled using different strategies and models. Generally, there is no clear consensus about which model is best for a task. Consequently, some form of trial and error can often not be avoided. This is supported by the “no free lunch” theorem stating that any algorithm outperforms any other algorithm on some datasets as well as by empirical studies

. The choice of models depends on many factors such as data (dimensionality, number of observations, structuredness), data mining objectives (need for best possible expected outcome, need to explain results), and cost (focus on minimum human effort to build or operate). From the perspective of green data mining, performance is assessed in terms of energy consumption for model training and model use, for instance, for making predictions. For the evaluation phase the main goal is to review all steps involved in the construction of the model, and to verify whether the final model meets the defined business objectives. If the best model meets the evaluation criteria, then it is deployed. Deployment ranges from fabricating a report presenting the findings in an easy-to-comprehend manner to implementing a long running system. Such a system might learn continuously while often performing a prediction task.

4. Principles of green data mining

Grounded in concepts and ideas from the literature on Green IT as well as data mining and its processes,

Table 1: Factors and methods related to green data mining

Factor	Subfactors	Methods for Green Data Mining
Project Objectives and Execution	Performance specification; Make, buy, share	Transfer Learning
Data	Quantity; Quality; Representation; Data acquisition method; Data storage	Sampling, Active Learning, Dimensionality Reduction, Compression, Change of Data Representation, Data Aggregation
Computation (Analysis)	Structuring of computation; Choice/Training of models; Training of models	Reuse of intermediate results; Approximate Models/Algorithms
IT Infrastructure	Hardware, e.g., CPU, Storage	

we identified factors determining the ecological footprint of data mining and we developed principles for reducing this footprint (Table 1, Figure 1).

Green IT discusses institutional perspectives , the role of users, including their behavior and beliefs when using IT-based systems as well as technical concerns . Topics include computational methods, their implementation in software , hardware components of computers , datacenters, cloud computing , parallel data processing (for big data) , as well as organizational and business aspects such as sustainable value chains, green oriented procurement, and adoption of Green IT . Loeser et al. discussed constructs and practices from Green IT (and IS) with respect to sourcing, operations, disposal, governance and end products.

Current literature on data mining , in particular data mining processes , does not explicitly discuss environmental concerns of data mining but touches upon aspects related to computational efficiency and storage such as data reduction and approximate algorithms.

Next, we describe principles of green data mining related to the different steps of the CRISP-DM process. We first elaborate on those principles that pertain to all stages of the process (principles 1-3 in Figure 1), before we then turn to those which only address specific stages (principles 4-8).

Principle #1: Identify and focus on the most energy consuming phases

To maximize the outcome of time invested into making data mining more environmentally friendly, the focus should be on the most energy consuming factors. This analysis can be performed by investigating the factors listed in Table 1 and analyzing each process step shown in Figure 1. Which process steps and factors dominate energy consumption depends on the goals and particularities of the data mining endeavor. Project objectives such as predictive accuracy or required confidence in the analysis are very likely to have a profound impact on energy consumption, since they often indirectly influence the choice of computational methods and data. For example, recent "deep learning" methods have outperformed other machine learning approaches for multiple classification tasks. A data scientist might turn to deep learning to meet certain project objectives, because it achieves state-of-the-art performance with respect to accuracy but, at the same time, requires lots of data and computation. Data preparation does often only require simple techniques, but it might be dominating in terms of energy consumption if complex computationally expensive methods are needed to extract features from the data that are used in later phases of the process. Deployment might be the dominating step if a system is built for continuous usage with large amounts of data. Still, deployment might contribute very little to the overall energy consumption compared to model selection, if the goal of the data mining project is to derive a report supporting a one-time decision.

Principle #2: Share and re-use data, models, frameworks and skills

A data scientist might control make-or-buy decisions. For example, for marketing purposes, she might choose to acquire data from social media

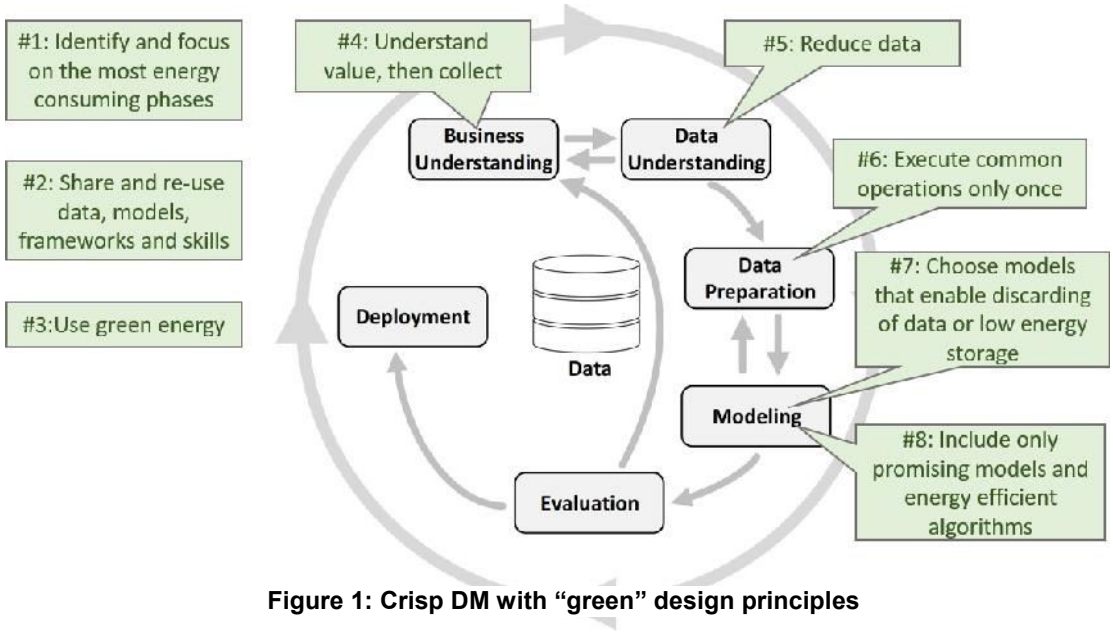


Figure 1: Crisp DM with "green" design principles

channels such as Twitter or Facebook and conduct the analysis by herself. She might also acquire models (implemented in software) to conduct the analysis. She might also decide to consult an external company to conduct the analysis or to obtain models. From an environmental perspective, outsourcing can be preferable if the contractor is more energy-efficient in extracting the demanded information, for instance, because of their prior experience and specialization, more energy efficient infrastructure, or even possession of relevant data. On a global scale, outsourcing of data analysis has the potential to involve less computation and to save energy.

Progress in the field of data science also relies on publicly available data, models, and development frameworks. Initiatives to make data available by research institutions and by governments help create entire ecosystems . State-of-the-art tools to develop (deep learning) models such as Google's Tensorflow are made freely available by large corporations. For such frameworks there are also numerous pre-trained models freely available, e.g., for image recognition based on the Imagenet dataset . Transfer learning is a technique that enables using knowledge from existing models trained for a specific task and dataset on different tasks . The idea is that some "knowledge" of a model can be transferred to another domain. Deep learning networks might benefit from reusing parameters or layers of an already trained network to reduce time (and energy consumption) on developing a new model. Thus, a green data scientist should also contribute data, models, and potentially extensions to frameworks to encourage re-use.

Principle #3: Use green energy

The use of renewable ("green") energy such as solar or wind should be maximized. Conceptually, the idea is to align computation with the availability of green energy. Technical realizations for data processing tasks for distributed data processing platforms (e.g., Hadoop) have been investigated. A system must predict the availability of green energy as well as brown energy and derive a schedule to maximize green energy use and to avoid using brown power at peak demand times. This strategy might also have a positive impact on energy costs as these increase with demand. The data scientist should identify the maximum possible slack in executing data processing tasks based on business objectives. More flexible scheduling allows for using more green energy.

Business understanding

The business understanding phase does typically not involve computation and as such generally does not contribute directly to the energy consumption. Still, understanding the business requirements and trends in the industry sector helps anticipate factors that influence energy consumption of later process steps, such as "What data are relevant and should be collected?" or "What precision of numbers is needed (over time)?" or "How frequently is a deployed system used?" or "How does the value of data change over time?"

Principle #4: Understand value, then collect and forget

Following the idea that "Data is the new oil"—a statement coined by Clive Humbly in 2006—it seems natural to collect as much data as possible, in particular given that storage is cheap and data might generate value "eventually." It is not uncommon that data can be obtained almost for free, for instance, in the form of trace data generated by users visiting a webpage. But, more data increases costs (due to storage and processing), requires more energy, impacts system performance and complexity and, additionally, enhances the risk of information overload. Query times to a database, for instance, increase with the amount of data stored in the database. The idea of collecting data only for the sake of collection has been criticized–"less data can be more value" . The data scientist should thus try to determine what data is relevant for the business or task at hand. Moreover, the quality of the data should be taken into consideration because data of inferior quality might require non-negligible effort for data cleaning .

Not all data has the same value. Even when data consists of a set of observations of the same kind, certain observations might be more valuable than others. For example, for observations, which should be split into classes, "difficult" to classify observations are often more helpful in training data mining models than "easy" to classify observations. Though computational methods can often determine the relevance of data with respect to well-defined metrics, a holistic understanding of the business, its objectives, data, and analytical methodology is essential to limit the collection of data. Leading data analytics companies such as Google embrace the idea of computing on more "little" data, that is, samples . This reasoning is well-founded not only based on statistical models, but also because models benefit from training data in a highly non-linear fashion with decreasing marginal gains given more data . Therefore, in some scenarios, reducing the volume of data might be feasible with considerable impact on energy consumption but only minor changes for other

relevant metrics. Since each model comes with its own strengths and weaknesses related to interpretability, robustness, speed of learning, etc., the overall assessment of advantages and disadvantages must be carefully conducted and aligned with underlying business objectives.

Data understanding

Principle #5: Reduce data

The data scientist might face the choice of what data to collect (or store). This choice must be made with great foresight in order not to miss any opportunity for data-driven value creation. Business understanding as well as an in depth understanding of the data are necessary. However, there are also multiple helpful techniques based on computational and statistical methods that might be supportive. We describe strategies to minimize the amount of data to be collected or used for training such as sampling and dimensionality reduction. These strategies can be employed to limit the number of attributes or observations, reducing precision and changing the representation of data.

Principle #5.1: Reduce number of data items

Often the data scientist can retrieve accurate results by looking at data samples or by using aggregated data. Data can also be categorized (or clustered) into groups, such that different attributes are relevant for some groups but not for others. A group might also be described using an average or median value. The grouping itself might be obtained by clustering algorithms, for instance, documents can be summarized using centroids obtained through clustering. Intuitively, one should maintain data that is most relevant to achieve a certain task. Active learning seeks to incrementally acquire relevant samples for learning. Thus, rather than having a passive model (or learner) that just uses the training data as given, an active learner might ask explicitly for data that is expected to yield maximal improvement in learning. Active learning is typically used in determining what data to collect. But the idea of active learning might also be used to assess the relevance of data and filter data accordingly. A model can be trained using active learning by incrementally adding the most important data items of the full dataset. The learning process might be stopped if there is no more data that improves the model beyond a small threshold. Unused data, which does not improve the model significantly, could then be discarded. Uncertainty sampling is the most prominent technique in active learning in the context of classification . It seeks to obtain labelled data, where there is most uncertainty about the correct class labels. Uncertainty sampling has been employed successfully for margin-based classifiers such as Support Vector Machines (SVMs) . Standard sampling techniques can also be helpful to reduce the amount of data. One of the simplest, but often sufficient approaches is to conduct simple random sampling—choosing each data point with the same probability without replacement of selected data points. In a case study on predicting conversion probabilities for two online retailers, Stange and Funk could show that only 1% of the data available to them was enough to achieve the optimal tradeoff between accuracy and the cost of collecting and processing the data. Stratified sampling is an appropriate sampling technique if groups are homogeneous, that is, data within groups has lower variance than data from distinct groups. One could also employ density-based sampling, for instance, assign samples with lower density a higher probability. This is useful if data from rare regions is highly important.

Principle #5.2: Reduce number or precision of attributes

The dataset might contain attributes that are irrelevant for the analysis. These attributes can be safely neglected. The relevance might depend on the type of data. For many text mining problems very frequent words—so-called stop words, such as “and”, ”the”, “is”, ”are”—can be ignored. In fact, removing unnecessary or noisy attributes such as stop words is often recommended . More generally, dimensionality reduction can be achieved by feature selection and extraction as well as type transformation . Feature selection techniques encompass filter and wrapper methods as well as their combination. Filter models assess the impact of features by some criterion independent of the model. Wrapper models train the model using a subset of features. An example of a filter model is the use of predictive attribute dependence, where the idea is that correlated features yield better outcomes than uncorrelated ones. Therefore, the relevance of an attribute might be determined by assessing the classification accuracy when using all other attributes to predict the attribute. These techniques can be employed to remove attributes that do not reach a minimum relevance threshold. Since many of the techniques are of heuristic nature, the impact of the removal of data that is deemed irrelevant should be tested, for instance, by comparing models being trained on the full and the reduced attribute set. Attribute reduction can also lead to an increase in accuracy, e.g., for decision trees

Feature extraction is often performed through axis rotations in a way that axes are sorted according to their ability to reconstruct data with minimal error.

Axes with negligible impact on data reconstruction can be removed. The derived dataset can often be used to train a model or it might be used to reconstruct the original data, which in turn is used for training. The prior approach is preferable, since a lesser volume of data must be processed. Prominent techniques include singular value decomposition (SVD), and a special case called principal component analysis (PCA).

SVD and similar techniques for feature extraction solve an optimization problem. This can be time consuming, making potential energy savings questionable. Random projections , where data is projected onto random manifolds, are a more simple and efficient dimensionality reduction technique. However, to achieve the same approximation guarantees more dimensions are needed than for SVD. Random projections preserve Euclidean distances according to the Johnson-Lindenstrauss Lemma as well as similarity computed using dot products, but random projections (as well as other dimensionality reduction techniques) do not preserve metrics such as the Manhattan distance. Therefore, some care is needed to ensure correct outcomes, when applying dimensionality reduction techniques. There is also empirical evidence comparing learning outcomes on the original data to outcomes on the data with reduced dimensionality. Unfortunately, the comparison neglects metrics relevant to energy, e.g., computation time.

Aggarwal describes dimensionality reduction with type transformation as the change of data from a more complex to a less complex type. For instance, graphs can be expressed as multidimensional data that might potentially be easier (and faster) to process. Time series can also be transformed to multi-dimensional data using the Haar Wavelet Transform or Fourier Transformation that both express the data using a (small) set of orthogonal functions. This form of data compression typically implies a loss of precision. Often, a dataset might only contain a few informative attributes and, therefore, the loss of precision might be very small, while achieving a substantial amount of data reduction. A high level understanding of the data mining task helps the data scientist choose a suitable dimensionality reduction technique. A technique might distort some instances more than others, and a small number of instances that are very different in the original context can be very similar in the space with reduced dimensions. For tasks like outlier detection this can be inacceptable, since outliers might be transformed so that they are not identifiable in the transformed data. Other tasks such as segmenting data into unspecified groups (clustering) might be less impacted by altering a few instances in a non-desirable way.

Principle #5.3: Change data representation

Data can be described in many ways without any loss of information, using lossless compression algorithms . This means that data is transformed among different representations without any effect on the minable knowledge. The green data scientist should prefer the representation that requires the least amount of storage, the least amount of computational effort to process throughout the data mining task, and the least amount of computation to create from the original data description.

A sequence of can be written more compactly as . Another form of encoding is difference encoding, where differences between two elements are stored, e.g. Difference encoding is often beneficial for time-series data, where commonly there is a strong dependency between consecutive data points. It is also possible to store only non-zero elements with indexes, e.g., the sequence 0,0,0,0,99,99 becomes 4:99, 5:99. In multiple dimensions such data structures are called sparse matrices. There are many applications where zero entries are common, e.g., document-term matrices representing textual documents and user-item matrices used to derive recommendations.

Numerous compression algorithms can be used to alter the data representation: General purpose algorithms such as Lempel-ziv as well as algorithms tailored to specific types of data. Sakr, for instance, surveys algorithms for XML data compressions. A dataset can be compressed in such a way that the entire dataset must be decompressed to access a single element. A compressed dataset might also allow for even faster access and manipulation of data than non-compressed data. For large matrices in a sparse matrix representation, for instance, some manipulations such as multiplication of two matrices are often faster. Compression and decompression also consume energy and, thus, data compression might or might not be beneficial depending on the number of required compress and decompress operations. General purpose algorithms allow to specify how much effort they should invest into finding the representation that minimizes space. Some algorithms take advantage of compressed representations and work on them directly, whereas others require an uncompressed representation. In case data is transferred across networks or is infrequently accessed, compression is even more appealing.

Principle #5.4: Accurate specification of attribute requirements

Whereas discrete attribute values stem from a fixed set of values, attributes with continuous values are

stored with a specific precision. The precision of individual attributes as well as the set of possible values can be defined by specifying an attribute type. For example, for an attribute containing temperature measurements, a data scientist might specify a precision of 0.001 degrees and a range of feasible values such as [0,100] as so called "domain constraint" in database systems . As a next step a data type can be chosen that meets these requirements and uses the least amount of storage—for instance, databases provide a set of data types according to the SQL standard, whereas programming languages usually follow the IEEE standards for floating point, integer, and other data types. The data type also determines the amount of storage and impacts the time and energy to conduct operations on data. The green data scientist should specify reasonable requirements. Choosing inappropriate types might more than double the amount of needed storage. For example, choosing an integer type (64 bits) rather than a (single) byte type (8 bits) for an array of many values leads to an increase of a factor of almost eight in memory demand.

Domain constraints depend on the data source, the range of the data, and the intended application: For sensor data, the accuracy is given by the maximum precision that seems achievable in the next years. For financial data, the needed accuracy might be given by the smallest unit, that is, one cent or one dollar. For time information, a precision up to milliseconds might not yield better outcomes than maintaining timestamps with hourly precision. For images, accuracy can be translated to the maximal resolution in terms of number of pixels or color depth that is beneficial for the analysis.

Data preparation and modeling

Principle #6: Execute common operations only once

Data preparation should be structured in such way that common preparation operations for multiple models are executed only once. For example, it can be reasonable to store a version of pre-processed data after general transformation and cleaning steps have been performed. The principle of factoring out common operations is already known, for instance, in the context of the Extract-Transform-Load (ETL) process optimization for data warehouses . The idea of storing temporary results has also been applied in the context of ETL processes and it is an integral part of the distributed data processing for Map-Reduce jobs. In both cases, the goal is fault tolerance rather than energy optimization. Strategies for identifying data processing results likely to be reused and thus worth storing have been investigated, too—for instance, for Map-Reduce jobs.

Principle #7: Choose models that enable discarding of data or low energy storage

Data lifecycle management has embraced the idea of moving data from high-cost to low-cost storage, for instance, moving data between storage tiers based on the value of data . Energy consumption and accessibility of stored data are typically negatively correlated: The easier it is to access data the more energy is required to maintain the data. Keeping data on a (magnetic) tape storage is much more energy efficient than keeping the same amount of data in the main memory of a computer. The former consumes energy only upon access, whereas main memory consumes energy even if no data is accessed. By her choices the data scientist determines the level of accessibility to data and thereby also the type of storage and amount of energy needed. The data scientist should thus be able to assess the relevance of data (over time) and assess the possibility to discard (older) data, compress (older) data, or work on summarized data. The availability of (old) data impacts the methodology that can be chosen, and the chosen methodology might also impact the data that must be stored. This is a key concern for long running systems, where data accumulates over time and models can be adjusted from time to time using newly available data. Some models can be trained incrementally using online learning algorithms, while others require the full dataset including all prior data, even in case only minor updates should be made due to new data using offline learning algorithms. For some models online as well as offline algorithms exist. Consider a system that classifies messages as spam or not spam. Such a system can be built by training a model based on previously classified messages. Since spammers adjust their strategies and style of messages, the system needs continuous updates—that is, learning. Whereas in an online learning scenario, data might be discarded after training the model, in the offline learning scenario it has to be kept.

Minimizing data access and thereby allowing to move data to energy friendly mediums is a viable option. But discarding data is a risky endeavor. What if the existing model should be replaced by a new model? Is it possible to change a model when all historic training data has been discarded? A careful assessment and management of risks is necessary. Various techniques from the domain of machine learning support reducing the need to keep data. One way is to use transfer learning by generating training data from the existing model for a new model, that is, to create labeled data in case unlabeled data is

available or can itself be generated. The disadvantage of this approach is that the generated labels are usually less accurate than the labels of the original dataset. Training data for the new model might still be highly beneficial despite transferred knowledge, but transfer learning can help reduce the amount of data needed to achieve good performance. Furthermore, training data can be enhanced by artificial training data that are a modification of existing data, thereby leading to improved results . Marginal returns decrease with additional data , and the impact in performance of having to retrain a new model might be small, even if just a small fraction of all data is retained.

Principle #8: Include only promising models and energy efficient algorithms

The traditional model selection process focuses almost exclusively on picking the model that yields the best results in terms of data mining-task-specific metrics such as accuracy or F-score for classification. A data scientist can base her model selection by comparing such metrics using empirical and theoretical comparisons (on similar datasets). The green data scientist, however, should also take into account energy consumption due to training, operating, and potentially data storage. Minor differences in task specific metrics might still be tolerable according to overall business objectives. It is not recommended to use all model and optimization algorithms as part of the computational selection process, because this leads to high energy costs. Ideally, the model candidates (and optimization algorithms) are limited to models that are likely to yield good results in terms of the desired metrics including energy efficiency. To this end, theoretical and empirical evidence should be leveraged.

A data scientist faces the choice of selecting model candidates and (hyper)parameter optimization algorithms. Energy costs are often determined by the effort to train and apply the model, that is, for predictions.

Principle #8.1: Leverage theoretical insights

Existing literature only gives limited advice on how to select the best methods for a dataset without trying them on the dataset at hand. Manning et al. advocate the use of high bias classifiers if little data is available. Properties of the learning algorithm are not the only factor impacting energy consumption. The number of hyperparameters and the effort to optimize these parameters also play a vital role. There are little theoretical foundations with respect to the best choice of hyperparameter optimization methods. The field is subject to current research . One theoretical insight is that obvious and intuitive techniques such as a systematic grid search might be inferior even to unstructured random search.

Models to describe the energy efficiency of systems and algorithms have been discussed from different perspectives such as power management , energy per low level operation (e.g., low level operations per Watt), or models involving hardware components such as CPUs and memory . However, none of these metrics seems suitable for quantifying the energy efficiency of models in the context of data mining. A data scientist usually works on a higher level of abstraction than individual hardware components and low-level CPU instructions that are the focus of many of these metrics. Theoretical computer science analyses algorithms in terms of running time. Running time, or time complexity, is the count of abstract, higher level operations needed to solve a task. The notion of time complexity can be applied to a single computer but also to a cluster of computers. In the field of parallel computing, one might simply aggregate the operations of all computers. This neglects costs due to information exchange between computers. Distributed systems such as clusters running data analytics frameworks such as Hadoop or Spark can also involve significant costs due to communication or idling (waiting for work). Generally, costs for communication, computation, and idling are tradeable . Many existing data mining algorithms are analyzed using the classical time complexity metric for a single computer, where the running time is often expressed as a function of the number of observations in a dataset and the number of dimensions. From the perspective of a green data scientist, algorithms with small time complexity seem preferable. But theoretical bounds might be coarse and, furthermore, often they neglect constants as part of the analysis process that might be of practical relevance. Therefore, empirical investigation might be more meaningful.

Principle #8.2: Leverage empirical knowledge

To the best of our knowledge a thorough comparison of learning algorithms for model parameters with respect to energy related concerns does not exist. Some works do provide empirical results for running-time of a few models, e.g., in the field of density based clustering. Running time seems to be a viable surrogate metric for measuring energy consumption of models for training and operation. For other metrics such as accuracy, multiple publications provide comparisons .

Hyperparameters often have a profound impact on model performance . To optimize hyperparameters multiple strategies exist. Some techniques try to reduce the time (and energy) for model selection by

training models on samples of data and predicting performance on the full dataset. Some optimization techniques allow to specify time constraints that guide the model selection process. Unfortunately, empirical comparisons do not report on the overall
energy consumption for training, but rather focus on other metrics such as accuracy.

5. Conclusion and future work

We introduced principles for green data mining based on the CRISP-DM methodology. Our principles apply to various phases of the process, impacting managerial decisions (e.g., make-or-buy) as well as technical questions (e.g., which model to use to conserve energy?). Creating a platform allowing to share information on model performance based on hyperparameter settings and datasets will not only be valuable for fellow data scientists, but also for
improving hyperparameter learning algorithms . Aside from empirical contributions, theoretical insights related to model selection could advance the field of green data mining. Furthermore, a detailed evaluation of the proposed principles can help in their application.

6. References

[1] Albers, S., "Energy efficient algorithms", Communications of the ACM, 53(5), 2010, pp. 86-96.
[2] Aggarwal, C. C., Data mining: The Textbook, 2015.
[3] Behrens, J. T., "Principles and procedures of exploratory data analysis", Psychological Methods, 1997.
[4] Bengio, Y., "Deep learning of representations for unsupervised and transfer learning", Proc. of ICML Workshop on Unsupervised and Transfer Learning, 2012.
[5] Bergstra, J. and Bengio, Y., "Random search for hyper- parameter optimization", Journal of Machine Learning Research, 13(Feb), 2012, pp. 281-305.
[6] Borra, S., and Di Ciaccio, A., "Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods", Computational statistics & data analysis, 54(12), 2010, pp. 2976-2989.
[7] Brooks, S., Wang, X., and Sarker, S., "Unpacking green IT: A review of the existing literature", In Proc. of the Americas Conf. on Information Systems (AMCIS), 2010.
[8] Capra, E., and Merlo, F., "Green IT: Everything starts from the software", In Proc. of European Conf. on Information Systems (ECIS), 2009.
[9] Caruana, R., and Niculescu-Mizil, A., "An empirical comparison of supervised learning algorithms", In Proc. of Int. Conf. on Machine learning, ACM, 2006.
[10] Cook, G., Pomeranz, D., Rohrbach, K., and Johnson, B., Clicking Clean: A Guide to Building the Green Internet. Greenpeace Inc., Washington, D.C, 2015.

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

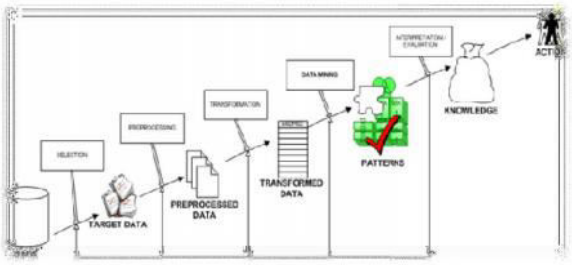
BOOSTER IN HIGH DIMENSIONAL DATA CLASSIFICATION

SANNIDHI SUBBA RAO

Dept of CSE *Malla Reddy College of Engineering* Maisammaguda, Hyderabad
sannidhisubbarao@gmail.com

Abstract— Classification problems in high dimensional data with a small number of observations are becoming more common especially in microarray data. During the last two decades, lots of efficient classification models and feature selection (FS) algorithms have been proposed for higher prediction accuracies. However, the result of an FS algorithm based on the prediction accuracy will be unstable over the variations in the training set, especially in high dimensional data. This paper proposes a new evaluation measure Q-statistic that incorporates the stability of the selected feature subset in addition to the prediction accuracy. Then, we propose the Booster of an FS algorithm that boosts the value of the Q- statistic of the algorithm applied. Empirical studies based on synthetic data and 14 microarray data sets show that Booster boosts not only the value of the Q-statistic but also the prediction accuracy of the algorithm applied unless the data set is intrinsically difficult to predict with the given algorithm.

INTRODUCTION



Structure of Data Mining

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases..

Data mining consists of five major elements:

1) Extract, transform, and load transaction data onto the data warehouse system.
2) Store and manage the data in a multidimensional database system.
3) Provide data access to business analysts and information technology professionals.
4) Analyze the data by application software.
5) Present the data in a useful format, such as a graph or table.

Different levels of analysis are available:

- Artificial neural networks: Non-linear predictive models that learn through training and resemble biological neural networks in structure.
- Genetic algorithms: Optimization techniques that use process such as genetic combination, mutation, and natural selection in a design based on the concepts of natural evolution.
- Decision trees: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID). CART and CHAID are decision tree techniques used for classification of a dataset. They provide a set of rules that you can apply to a new (unclassified) dataset to predict which records will have a given outcome. CART segments a dataset by creating 2-way splits while CHAID segments using chi square tests to create multi-way splits. CART typically requires less data preparation than CHAID.
- Nearest neighbor method: A technique that classifies each record in a dataset based on a combination of the classes of the *k* record(s) most similar to it in a historical

dataset (where k=1). Sometimes called the k-nearest neighbor technique.

- Rule induction: The extraction of useful if-then rules from data based on statistical significance.
- Data visualization: The visual interpretation of complex relationships in multidimensional data. Graphics tools are used to illustrate data relationships.

LITERATURE SURVEY

Biomarker discovery is an important topic in biomedical applications of computational biology, including applications such as gene and SNP selection from high-dimensional data. However, robustness of biomarkers is an important issue, as it may greatly influence subsequent biological validations. First contribution is a general framework for the analysis of the robustness of a biomarker selection algorithm. Secondly, they conducted a large-scale analysis of the recently introduced concept of ensemble feature selection, where multiple feature selections are combined in order to increase the robustness of the final set of selected features. We focus on selection methods that are embedded in the estimation of support vector machines (SVMs). SVMs are powerful classification models that have shown state-of-the-art performance on several diagnosis and prognosis tasks on biological data. Their feature selection extensions also offered good results for gene selection tasks. they show that the robustness of SVMs for biomarker discovery can be substantially increased by using ensemble feature selection techniques, while at the same time improving upon classification performances. The proposed methodology is evaluated on four microarray datasets showing increases of up to almost 30% in robustness of the selected biomarkers, along with an improvement of ~15% in classification performance. The stability improvement with ensemble methods is particularly noticeable for small signature sizes (a few tens of genes), which is most relevant for the design of a diagnosis or prognosis model from a gene signature [1].

Storing and using specific instances improves the performance of several supervised learning algorithms. These include algorithms that learn decision trees, classification rules, and distributed networks. However, no investigation has analyzed algorithms that use only specific instances to solve incremental learning tasks. In this paper, we describe a framework and methodology, called instance-based learning, that generates classification predictions using only specific instances. Instance-based learning algorithms do not maintain a set of abstractions derived from specific instances. This approach extends the nearest neighbor algorithm, which has large storage requirements. We describe how storage requirements can be significantly reduced with, at most, minor sacrifices in learning rate and classification accuracy. While the storage-reducing algorithm performs well on several real-world databases, its performance degrades rapidly with the level of attribute noise in training instances. Therefore, we extended it with a significance test to distinguish noisy instances. This extended algorithm's performance degrades gracefully with increasing noise levels and compares favorably with a noise-tolerant decision tree algorithm [2].

Diffuse large B-cell lymphoma (DLBCL), the most common subtype of non-Hodgkin's lymphoma, is clinically heterogeneous: 40% of patients respond well to current therapy and have prolonged survival, whereas the remainder succumb to the disease. [3] proposed that this variability in natural history reflects unrecognized molecular heterogeneity in the tumours. Using DNA microarrays, They have conducted a systematic characterization of gene expression in B-cell malignancies and show that there is diversity in gene expression among the tumours of DLBCL patients, apparently reflecting the variation in tumour proliferation rate, host response and differentiation state of the tumour. They identified two molecularly distinct forms of DLBCL which had gene expression patterns indicative of different stages of B-cell differentiation. One type expressed genes characteristic of germinal centre B cells ('germinal centre B-like DLBCL'); the second type expressed genes normally induced during in vitro activation of peripheral blood B cells ('activated B-like DLBCL'). Patients with germinal centre B-like DLBCL had a significantly better overall survival than those with activated B-like DLBCL. The molecular classification of tumours on the basis of gene

expression can thus identify previously undetected and clinically significant subtypes of cancer.

Oligonucleotide arrays can provide a broad picture of the state of the cell, by monitoring the expression level of thousands of genes at the same time. It is of interest to develop techniques for extracting useful information from the resulting data sets. [4] report the application of a two-way clustering method for analyzing a data set consisting of the expression patterns of different cell types. Gene expression in 40 tumor and 22 normal colon tissue samples was analyzed with an Affymetrix oligonucleotide array complementary to more than 6,500 human genes. An efficient two-way clustering algorithm was applied to both the genes and the tissues, revealing broad coherent patterns that suggest a high degree of organization underlying gene expression in these tissues. Coregulated families of genes clustered together, as demonstrated for the ribosomal proteins. Clustering also separated cancerous from noncancerous tissue and cell lines from in vivo tissues on the basis of subtle distributed patterns of genes even when expression of individual genes varied only slightly between the tissues. Two-way clustering thus may be of use both in classifying genes into functional groups and in classifying tissues based on gene expression.

Early detection of ventricular fibrillation (VF) is crucial for the success of the defibrillation therapy in automatic devices. A high number of detectors have been proposed in [5] based on temporal, spectral, and time–frequency parameters extracted from the surface electrocardiogram (ECG), showing always a limited performance. The combination ECG parameters on different domain (time, frequency, and time–frequency) using machine learning algorithms has been used to improve detection efficiency. However, the potential utilization of a wide number of parameters benefiting machine learning schemes has raised the need of efficient feature selection (FS) procedures. In this study, we propose a novel FS algorithm based on support vector machines (SVM) classifiers and bootstrap resampling (BR) techniques. We define a backward FS procedure that relies on evaluating changes in SVM performance when removing features from the input space. This evaluation is achieved according to a nonparametric statistic based on BR. After simulation studies, we benchmark the performance of our FS algorithm in AHA and MIT-BIH ECG databases. Our results show that the proposed FS algorithm outperforms the recursive feature elimination method in synthetic examples, and that the VF detector performance improves with the reduced feature set.

IMPLEMENTATION

MODULES:

- Dataset Collection
- Feature Selection
- Removing Irrelevant Features
- Booster accuracy

MODULES DESCSRIPTION:

Dataset Collection:

To collect and/or retrieve data about activities, results, context and other factors. It is important to consider the type of information it want to gather from your participants and the ways you will analyze that information. The data set corresponds to the contents of a single database table, or a single statistical data matrix, where every column of the table represents a particular variable. after collecting the data to store the Database.

Feature Selection:

This is a hybrid measure of the prediction accuracy of the classifier and the stability of the selected features. Then the paper proposes Booster on the selection of feature subset from a given FS algorithm. The data sets from original data set by re sampling on sample space. Then FS algorithm is applied to each of these re sampled data sets to obtain different feature subsets. The union of these selected subsets will be the feature subset obtained by the Booster of FS algorithm. the Booster of an algorithm boosts not only the value of Q-statistic but also the prediction accuracy of the classifier applied. Several studies based on re sampling technique have been done to generate different data sets for classification problem , and some of the studies utilize re sampling on the feature space . The purposes of all these

studies are on the prediction accuracy of classification without consideration on the stability of the selected feature subset. FS algorithms— FAST, FCBF, and mRMR—and their corresponding Boosters, we apply k-fold cross validation. For this, k training sets and their corresponding k test sets are generated. For each training set, Booster is applied to obtain V . Classification is performed based on the training set with the selection V , and the test set is used for prediction accuracy

Removing Irrelevant Features:

The features of high dimensional microarray data are irrelevant to the target feature and the proportion of relevant features or the percentage of up-regulated Finding relevant features simplifies learning process and increases prediction accuracy. The finding, however, should be relatively robust to the variations in training data, especially in biomedical study, since domain experts will invest considerable time and efforts on this small set of selected features. The pre-processing steps to find weakly relevant features based on t-test and to remove irrelevant features based on MI.FS in high dimensional data needs preprocessing process to select only relevant features or to filter out irrelevant features. the selected subsets V1; . . . ; Vb obtained by s consist only of the relevant features where redundancies are removed, V will include more relevant features where redundancies are removed. Hence, V will induce smaller error of selecting irrelevant features. However, if s does not completely remove redundancies, V may result in the accumulation of larger size of redundant features. find more relevant features but may include more irrelevant features, and also may induce more redundant features. This is because no FS algorithm can select all relevant features while removing all irrelevant features and redundant features.

Booster accuracy:

The Booster of an FS algorithm that boosts the value of the Q-statistic of the algorithm applied. Empirical studies based on synthetic data Empirical studies show that the Booster of an algorithm boosts not only the value of Q-statistic but also the prediction accuracy of the classifier applied. Booster is simply a union of feature subsets obtained by a resembling technique. The resembling is done on the sample space. Booster needs an FS algorithm s and the number of partitions b. When s and b are needed to be specified, we will use notation s-Booster. Hence, s-Booster1 is equal to s since no partitioning is done in this case and the whole data is used. When s selects relevant features while removing redundancies, s-Booster will also select relevant features while removing redundancies. the notation FAST-Booster, FCBF-Booster, and mRMR-Booster for the Booster of the corresponding FS algorithm. we will evaluate the relative performance efficiency of s-Booster over the original FS algorithm s based on the prediction accuracy and Q-statistic.two Boosters, FAST-Booster , FCBF-Booster and mRMR-Booster. mRMR- Booster improves accuracy considerably: overall average accuracy. One interesting point to note here is that mRMR-Booster is more efficient in boosting the accuracy .we can observe that FAST-Booster also improves accuracy, but not as high as mRMR

SYSTEM ANALYSIS

EXISTING SYSTEM:

One often used approach is to first discretize the continuous features in the preprocessing step and use mutual information (MI) to select relevant features. This is because finding relevant features based on the discretized MI is relatively simple while finding relevant features directly from a huge number of the features with continuous values using the definition of relevancy is quite a formidable task.

Several studies based on resampling technique have been done to generate different data sets for classification problem and some of the studies utilize resampling on the feature space.

The purposes of all these studies are on the prediction accuracy of classification without consideration on the stability of the selected feature subset.

DISADVANTAGES OF EXISTING SYSTEM:

Most of the successful FS algorithms in high dimensional problems have utilized forward selection method but not considered backward elimination method

since it is impractical to implement backward elimination process with huge number of features.

A serious intrinsic problem with forward selection is, however, a flip in the decision of the initial feature may lead to a completely different feature subset and hence the stability of the selected feature set will be very low although the selection may yield very high accuracy. Devising an efficient method to obtain a more stable feature subset with high accuracy is a challenging area of research.

PROPOSED SYSTEM:

This paper proposes Q-statistic to evaluate the performance of an FS algorithm with a classifier. This is a hybrid measure of the prediction accuracy of the classifier and the stability of the selected features. Then the paper proposes Booster on the selection of feature subset from a given FS algorithm.

The basic idea of Booster is to obtain several data sets from original data set by resampling on sample space. Then FS algorithm is applied to each of these resampled data sets to obtain different feature subsets. The union of these selected subsets will be the feature subset obtained by the Booster of FS algorithm.

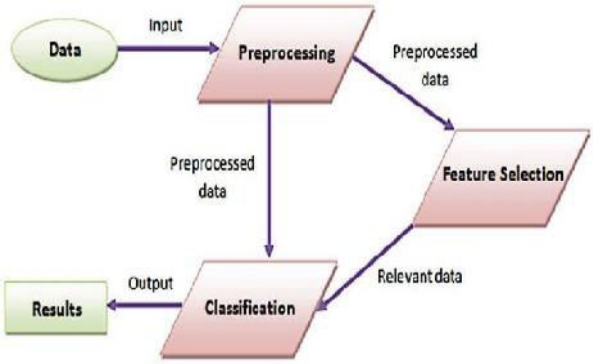
ADVANTAGES OF PROPOSED SYSTEM:

Empirical studies show that the Booster of an algorithm boosts not only the value of Q-statistic but also the prediction accuracy of the classifier applied.

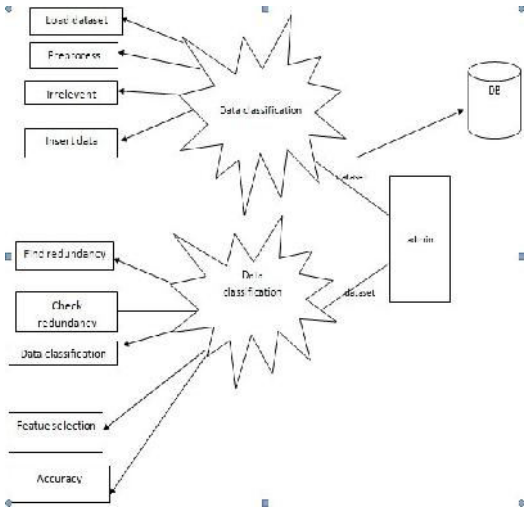
- We have noted that the classification methods applied to Booster do not have much impact on prediction accuracy and Q-statistic. Especially, the performance of mRMR-Booster was shown to be outstanding both in the improvements of prediction accuracy and Q-statistic.

SYSTEM DESIGN

SYSTEM ARCHITECTURE:



BLOCK DIAGRAM:

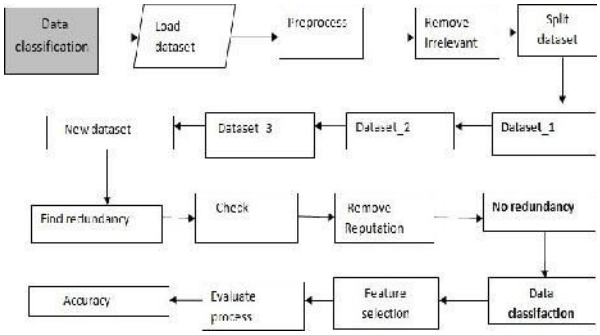


DATA FLOW DIAGRAM:

1. The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
2. The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
3. DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts

information flow and the transformations that are applied as data moves from input to output.

4. DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.



SYSTEM STUDY

FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

ECONOMICAL FEASIBILITY

TECHNICAL FEASIBILITY

SOCIAL FEASIBILITY

ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

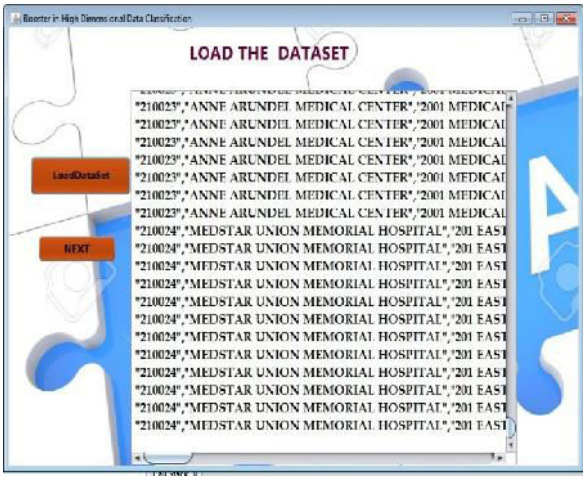
TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

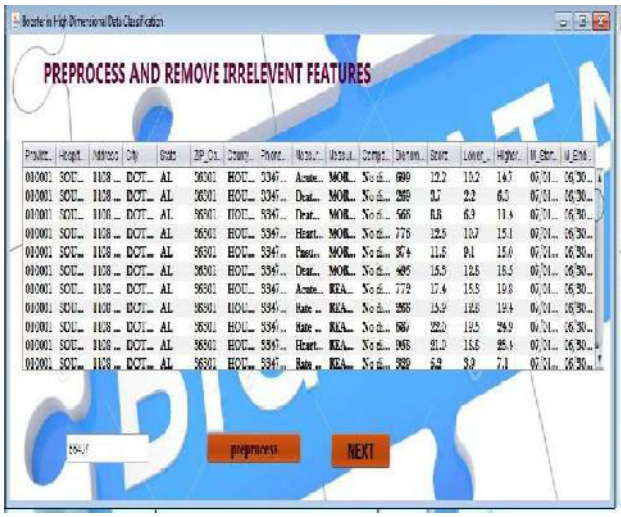
SOCIAL FEASIBILITY

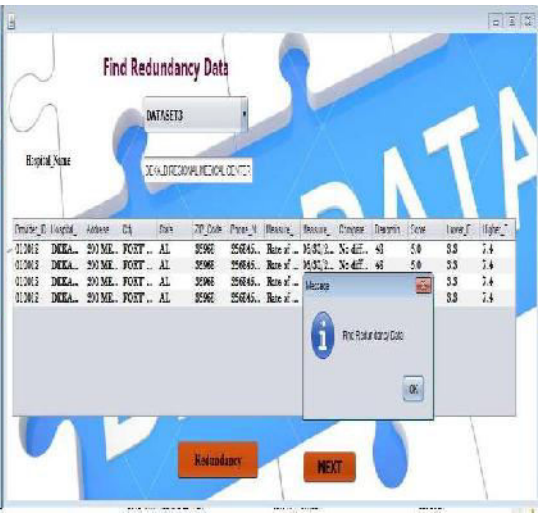
The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

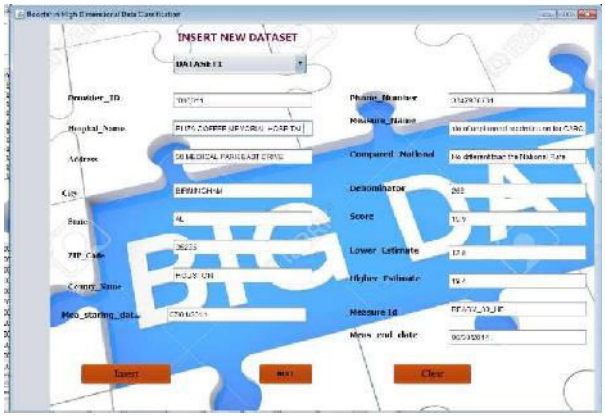
SCREEN SHOTS

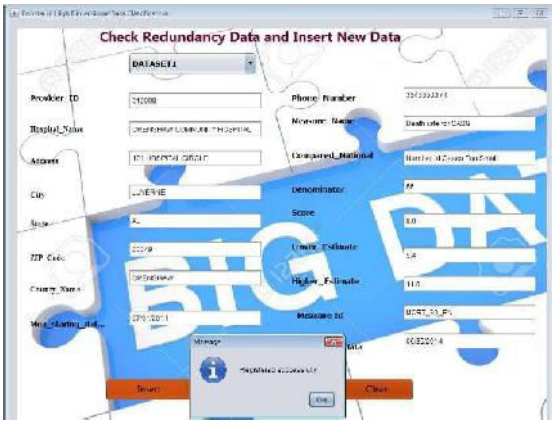


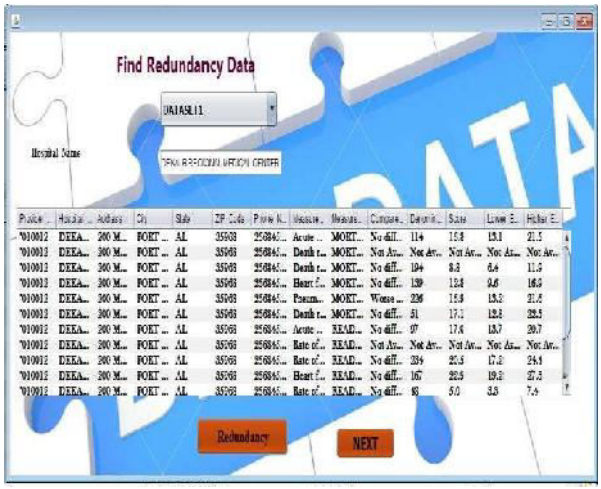
ISBN 978-93-88808-61-3

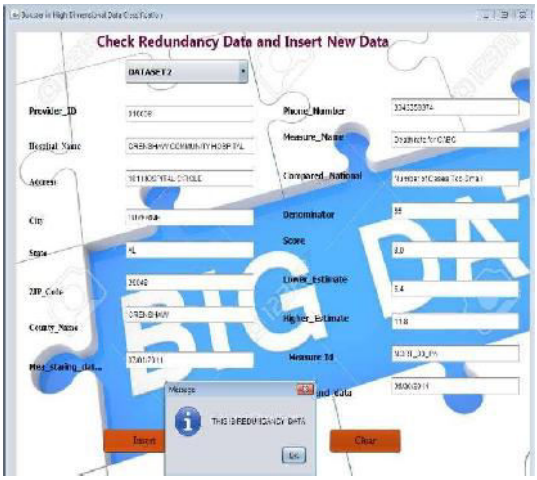




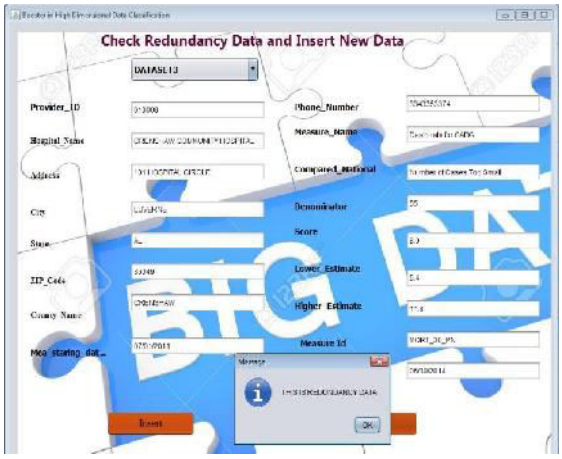


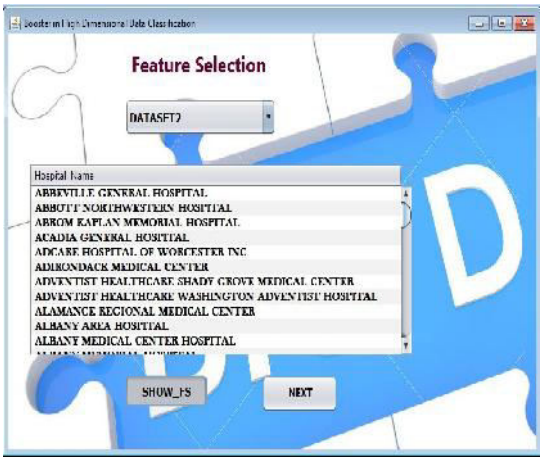


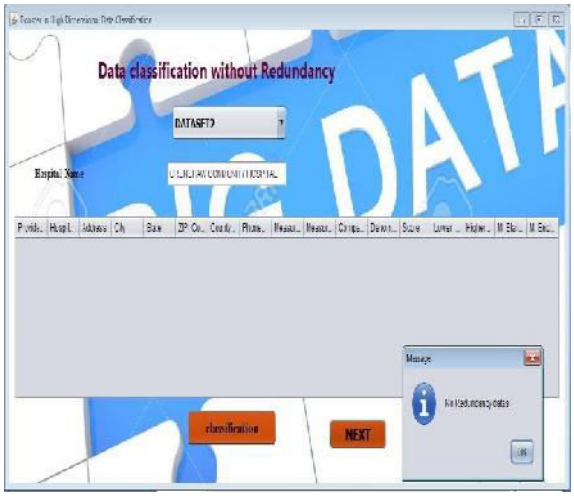


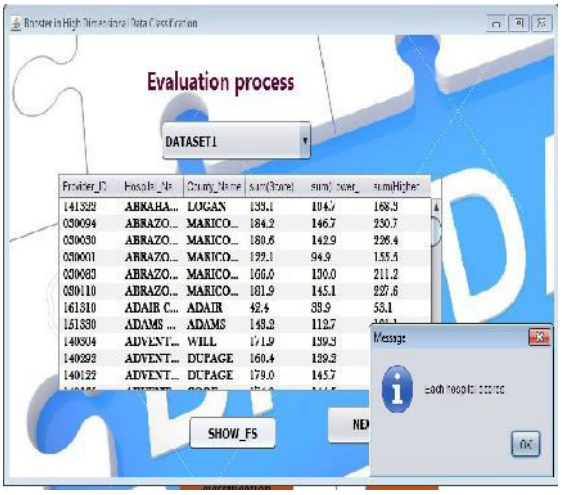


ISBN 978-93-88808-61-3

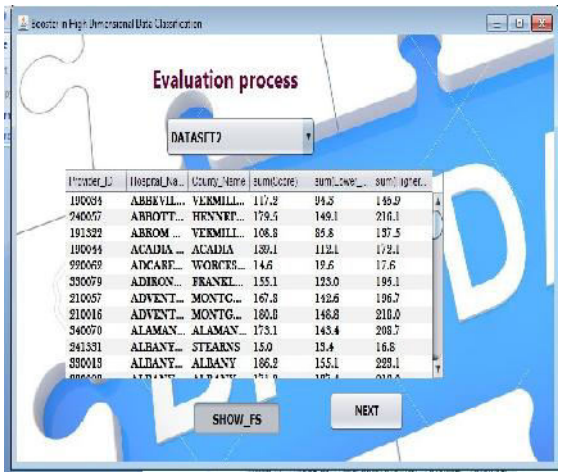


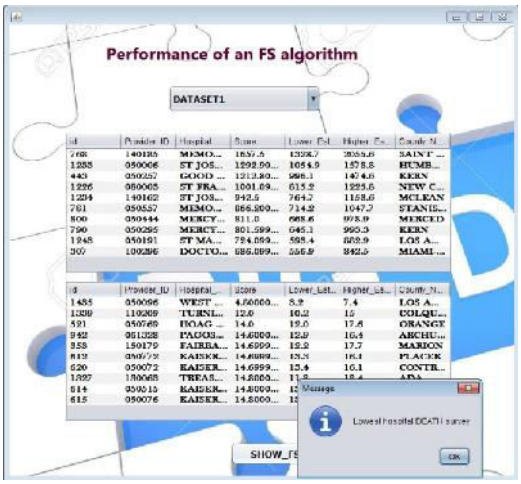


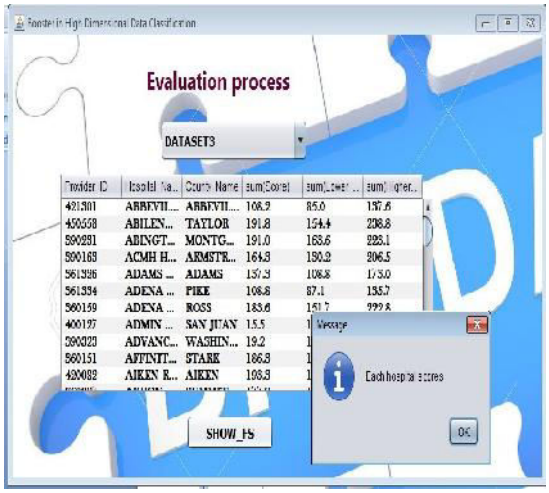


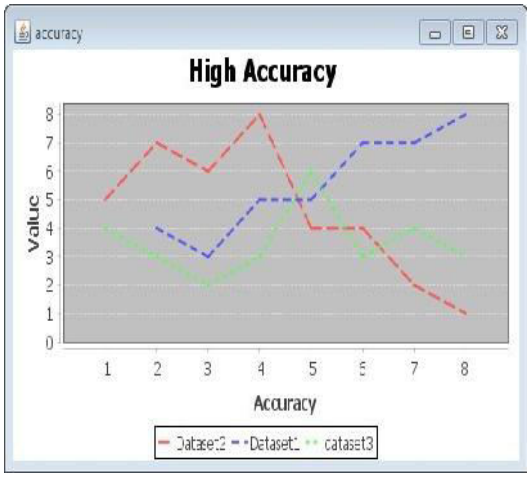


ISBN 978-93-88808-61-3

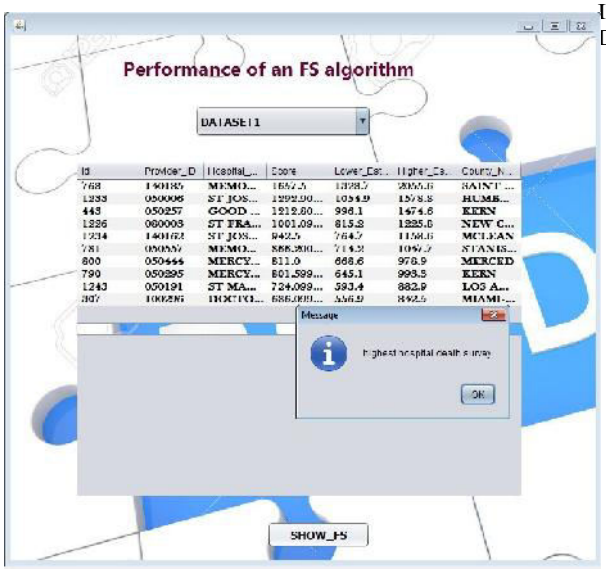








INPUT DESIGN AND OUTPUT DESIGN



INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

What data should be given as input?

How the data should be arranged or coded?

The dialog to guide the operating personnel in providing input.

Methods for preparing input validations and steps to follow when error occur.

OBJECTIVES

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.
2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.
3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow

OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.
2. Select methods for presenting information.
3. Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

Convey information about past activities, current status or projections of the

Future.

Signal important events, opportunities, problems, or warnings.

Trigger an action.

Confirm an action.

CONCLUSION

This paper proposed a measure Q-statistic that evaluates the performance of an FS algorithm. Q-statistic accounts both for the stability of selected feature subset and the prediction accuracy. The paper proposed Booster to boost the performance of an existing FS algorithm. Experimentation with synthetic data and 14 microarray data sets has shown that the suggested Booster improves the prediction accuracy and the Q-statistic of the three well-known FS algorithms: FAST, FCBF, and mRMR. Also we have noted that the classification methods applied to Booster do not have much impact on prediction accuracy and Q-statistic. Especially, the performance of mRMR-Booster was shown to be outstanding both in the improvements of prediction accuracy and Q-statistic.

It was observed that if an FS algorithm is efficient but could not obtain high performance in the accuracy or the Q-statistic for some specific data, Booster of the FS algorithm will boost the performance. However, if an FS algorithm itself is not efficient, Booster may not be able to obtain high performance. The performance of Booster depends on the performance of the FS algorithm applied. If Booster does not provide high performance, it implies two possibilities: the

data set is intrinsically difficult to predict or the FS algorithm applied is not efficient with the specific data set. Hence, Booster can also be used as a criterion to evaluate the performance of an FS algorithm or to evaluate the difficulty of a data set for classification.

REFERENCES

[1] T. Abeel, T. Helleputte, Y. V. de Peer, P. Dupont, and Y. Saeys, "Robust biomarker identification for cancer diagnosis with ensemble feature selection methods," Bioinformatics, vol. 26, no. 3, pp. 392–398, 2010.

[2] D. Aha and D. Kibler, "Instance-based learning algorithms," Mach. Learn., vol. 6, no. 1, pp. 37–66, 1991.

[3] A. A. Alizadeh, M. B. Eisen, R. E. Davis, C. M. Izidore, S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Tran, X. Yu, J. I. Powell, L. Yang, G. E. Marti, T. Moore, J. H. Jr, L. Lu, D. B. Lewis, R. Tibshirani, G. Sherlock, W. C. Chan, T. C. Greiner, D.
D. Weisenburger, J. O. Armitage, R. Warnke, R. Levy, W. Wilson, M. R. Grever, J. C. Byrd, D. Botstein, P. O. Brown, and L. M. Staudt, "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling," Nature, vol. 403, no. 6769, pp. 503–511, 2000.

[4] U. Alon, N. Barkai, D. A. Notterman, K. Gish, S. Ybarra, D. Mack, and A. J. Levine, "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays," Proc. Nat. Acad. Sci., vol. 96, no. 12, pp. 6745–6750, 1999.

[5] F. Alonso-Atienza, J. L. Rojo-Alvare, A. Rosado-Mu~noz, J. J. Vinagre, A. Garcia-Alberola, and G. Camps-Valls, "Feature selection using support vector machines and bootstrap methods for ventricular fibrillation detection," Expert Syst. Appl., vol. 39, no. 2, pp. 1956–1967, 2012.

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

Trends of Internet of Things in India

E Srinath,
Assistant Professor,
Dept. of CSE,
eslavath.sri77@gmail.com
Malla Reddy College of Engineering

Ch.Malleswar Rao
Assistant Professor
Dept. of CSE,
malleswar.538@gmail.com
Malla Reddy College of Engineering

Abstract: Rural India is the base of our Country. In Rural India, Harvesting and Agriculture is top bread-winning activity. In India about 70% of population depends upon farming and one third of the nation's capital comes from farming. Issues concerning agriculture have been always hindering the development of the country. The only solution to this problem is smart agriculture by modernizing the current traditional methods of agriculture. Hence the project aims at making agriculture smart using automation and IoT technologies. The highlighting features of this project includes smart GPS based remote controlled robot to perform tasks like weeding, spraying, moisture sensing, bird and animal scaring, keeping vigilance, etc. Secondly it includes smart irrigation with smart control and intelligent decision making based on accurate real time field data. Thirdly, smart warehouse management which includes temperature maintenance, humidity maintenance and theft detection in the warehouse. Controlling of all these operations will be through any remote smart device or computer connected to Internet and the operations will be performed by interfacing sensors, Wi-Fi or ZigBee modules, camera and actuators with micro-controller and raspberry pi.

Keywords: IoT, automation, Wi-Fi

I. INTRODUCTION

Agriculture is considered as the basis of life for the human species as it is the main source of food grains and other raw materials. It plays vital role in the growth of country's economy. It also provides large ample employment opportunities to the people. Growth in agricultural sector is necessary for the development of economic condition of the country. Unfortunately, many farmers still use the traditional methods of farming which results in low yielding of crops and fruits. But wherever automation had been implemented and human beings had been replaced by automatic machineries, the yield has been improved. Hence there is need to implement modern science and technology in the agriculture sector for increasing the yield. Most of the papers signifies the use of wireless sensor network which collects the data from different types of sensors and then send it to main server using wireless protocol. The collected data provides the information about different environmental factors which in turns helps to monitor the system. Monitoring environmental factors is not enough and complete solution to improve the yield of the crops. There are number of other factors that affect the productivity to great extent. These factors include attack of insects and pests which can be controlled by spraying the crop with proper insecticide and pesticides. Secondly, attack of wild animals and birds when the crop grows up. There is also possibility of thefts when crop is at the stage of harvesting. Even after harvesting, farmers also face problems in storage of harvested crop. So, in order to provide solutions to all such problems, it is necessary to develop integrated system which will take care of all factors affecting the productivity in every stages like; cultivation, harvesting and post harvesting storage. This paper therefore proposes a system which is useful in monitoring the field data as well as controlling the field operations which provides the Flexibility. The paper aims at making agriculture smart using automation and IoT technologies. The highlighting features of this paper includes smart GPS based remote controlled robot to perform tasks like; weeding, spraying, moisture sensing, bird and animal scaring, keeping vigilance, etc. Secondly, it includes smart irrigation with smart control based on real time field data. Thirdly, smart warehouse management which includes; temperature maintenance, humidity maintenance and theft detection in the warehouse. Controlling of all these operations will be through any remote smart device or computer connected to Internet and the operations will be performed by interfacing sensors, Wi-Fi or ZigBee modules, camera and actuators with micro-controller and raspberry pi.

II. LITERATURE REVIEW

The newer scenario of decreasing water tables, drying up of rivers and tanks, unpredictable environment present an urgent need of proper utilization of water. To cope up with this use of temperature and moisture sensor at suitable locations for monitoring of crops is implemented in. [1] An algorithm developed with threshold values of temperature and soil moisture can be programmed into a microcontroller-based gateway to control water quantity. The system can be powered by photovoltaic panels and can have a duplex communication link based on a cellular-Internet interface that allows data inspection and irrigation scheduling to be programmed through a web page. [2]
The technological development in Wireless Sensor Networks made it possible to use in monitoring and control of greenhouse parameter in precision agriculture.
[3] After the research in the agricultural field, researchers found that the yield of agriculture is decreasing day by day. However, use of technology in the field of agriculture

plays important role in increasing the production as well as in reducing the extra man power efforts. Some of the research attempts are done for betterment of farmers which provides the systems that use technologies helpful for increasing the agricultural yield.

A remote sensing and control irrigation system using distributed wireless sensor network aiming for variable rate irrigation, real time in field sensing, controlling of a site specific precision linear move irrigation system to maximize the productivity with minimal use of water was developed by Y. Kim . The system described details about the design and instrumentation of variable rate irrigation, wireless sensor network and real time in field sensing and control by using appropriate software. The whole system was developed using five in field sensor stations which collects the data and send it to the base station using global positioning system (GPS) where necessary action was taken for controlling irrigation according to the database available with the system. The system provides a promising low cost wireless solution as well as remote controlling for precision irrigation. [4]

In the studies related to wireless sensor network, researchers measured soil related parameters such as temperature and humidity. Sensors were placed below the soil which communicates with relay nodes by the use of effective communication protocol providing very low duty cycle and hence increasing the life time of soil monitoring system. The system was developed using microcontroller, universal asynchronous receiver transmitter (UART) interface and sensors while the transmission was done by hourly sampling and buffering the data, transmit it and then checking the status messages. The drawbacks of the system were its cost and deployment of sensor under thesoil which causes attenuation of radio frequency (RF) signals. [5]

III. SYSTEM OVERVIEW

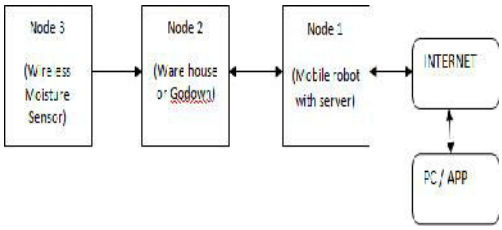


Figure 1: System overview

The paper consist of four sections; node1, node2, node3 and PC or mobile app to control system. In the present system, every node is integration with different sensors and devices and they are interconnected to one central server via wireless communication modules. The server sends and receives information from user end using internet connectivity. There are two modes of operation of the system; auto mode and manual mode. In auto mode system takes its own decisions and controls the installed devices whereas in manual mode user can control the operations of system using android app or PC commands.

IV. ARCHITECTURE OF THE SYSTEM

Node 1:

Node1 is GPS based mobile robot which can be controlled remotely using computer as well as it can be programmed so as to navigate autonomously within the boundary of field using the co-ordinates given by GPS module.

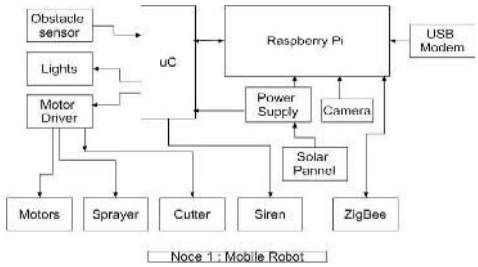


Figure 2: Node 1

The Remote controlled robot have various sensors and devices like camera, obstacle sensor, siren, cutter, sprayer and using them it will perform tasks like; Keeping vigilance, Bird and animal scaring, Weeding, and Spraying

Node 2:

Node2 will be the warehouse. It consists of motion detector, light sensor, humidity sensor, temperature sensor, room heater, cooling fan altogether interfaced with AVR microcontroller. Motion detector will detect the motion in the room when security mode will be ON and on detection of motion, it will send the alert signal to user via Raspberry pi and thus providing theft detection.

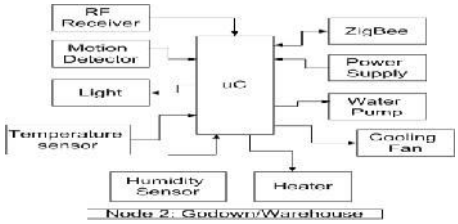


Figure 3: Node 2

Temperature sensor and Humidity sensor senses the temperature and humidity respectively and if the value crosses the threshold then room heater or cooling fan will be switched ON/OFF automatically providing temperature and humidity maintenance.Node2 will also controls water pump depending upon the soil moisture data sent by node3.

Node 3:

Node3 is a smart irrigation node with features like ; Smart control of water pump based on real time field data i.e. automatically turning on/off the pump after attaining the required soil moisture level in auto mode, Switching water pump on/off remotely via mobile or computer in manual mode, and continuous monitoring of soil moisture.

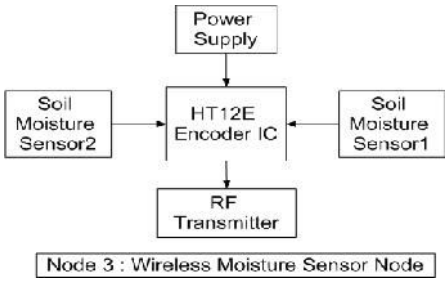


Figure 4: Node 3

In node3, moisture sensor transmits the data using HT12E Encoder IC and a RF transmitter. The transmitted data is received by node2 and there it is processed by microcontroller in order to control the operation of water pump.

Hardware used:

a) AVR Microcontroller Atmega 16/32:

The microcontroller used is, Low-power AVR® 8-bit Microcontroller, having 8K Bytes of In-System Self-programmable Flash program memory, Programmable Serial USART, 8-channel, 10-bit ADC, 23 Programmable I/O Lines.

b) ZigBee Module:

ZigBee is used for achieving wireless communication between Node1 and Node2. The range for Zigbee is roughly 50 meters and it can be increased using high power modules or by using network of modules. It operates on 2.4 GHz frequency. Its power consumption is very low and it is less expensive as compared to other wireless modules like Wi-Fi or Bluetooth. It is usually used to establish wireless local area networks.

c) Temperature Sensor LM35:

The LM35 is precision IC temperature sensor. Output voltage of LM35 is directly proportional to the Centigrade/Celsius of temperature. The LM35 does not need external calibration or trimming to provide accurate temperature range. It is very low cost sensor. It has low output impedance and linear output. The operating temperature range for LM35 is -55° to $+150^{\circ}$C. With rise in temperature, the output voltage of the sensor increases linearly and the value of voltage is given to the microcontroller which is multiplied by the conversion factor in order to give the value of actual temperature.

d) Moisture sensor:

Soil moisture sensor measures the water content in soil. It uses the property of the electrical resistance of the soil. The relationship among the measured property and soil moisture is calibrated and it may vary depending on environmental factors such as temperature, soil type, or electric conductivity. Here, It is used to sense the moisture in field and transfer it to microcontroller in order to take controlling action of switching water pump ON/OFF.

Humidity sensor:

The DHT11 is a basic, low-cost digital temperature and humidity sensor. It gives out digital value and hence there is no need to use conversion algorithm at ADC of the microcontroller and hence we can give its output directly to data pin instead of ADC. It has a capacitive sensor for measuring humidity. The only real shortcoming of this sensor is that one can only get new data from it only after every 2 seconds.

e) Obstacle sensor (Ultra-Sonic):

The ultra-sonic sensor operates on the principle of sound waves and their reflection property. It has two parts; ultra-sonic transmitter and ultra-sonic receiver. Transmitter transmits the 40 KHz sound wave and receiver receives the reflected 40 KHz wave and on its reception, it sends the electrical signal to the microcontroller. The speed of sound in air is already known.

Hence from time required to receive back the transmitted sound wave, the distance of obstacle is calculated. Here, it is used for obstacle detection in case of mobile robot and as a motion detector in ware house for preventing thefts. The ultra-sonic sensor enables the robot to detect and avoid obstacles and also to measure the distance from the obstacle. The range of operation of ultra-sonic sensor is 10 cm to 30 cm.

f) Raspberry Pi :

The Raspberry Pi is small pocket size computer used to do small computing and networking operations. It is the main element in the field of internet of things. It provides access to the internet and hence the connection of automation system with remote location controlling device becomes possible. Raspberry Pi is available in various versions. Here, model Pi 2 model B is used and it has quad-core ARM Cortex-A53 CPU of 900 MHz, and RAM of 1GB. it also has: 40 GPIO pins, Full HDMI port, 4 USB ports, Ethernet port, 3.5mm audio jack, video Camera interface (CSI), the Display interface (DSI), and Micro SD card slot.

Software's used:

a) AVR Studio Version 4:

It is used to write, build, compile and debug the embedded c program codes which are needed to be burned in the microcontroller in order to perform desired operations. This software directly provides .hex file which can be easily burned into the microcontroller.

b) Proteus 8 Simulator:

Proteus 8 is one of the best simulation software for various circuit designs of microcontroller. It has almost all microcontrollers and electronic components readily available in it and hence it is widely used simulator.

It can be used to test programs and embedded designs for electronics before actual hardware testing. The simulation of programming of microcontroller can also be done in Proteus. Simulation avoids the risk of damaging hardware due to wrong design.

c) Dip Trace:
Dip race is EDA/CAD software for creating schematic diagrams and printed circuit boards. The developers provide multi-lingual interface and tutorials (currently available in English and 21 other languages). Dip Trace has 4 modules: Schematic Capture Editor, PCB Layout Editor with built-in shape-based auto router and 3D Preview & Export, Component Editor, and Pattern Editor.

d) SinaProg:
SinaProg is a Hex downloader application with AVR Dude and Fuse Bit Calculator. This is used to download code/program and to set fuse bits of all AVR based microcontrollers.

e) Raspbian Operating System:
Raspbian operating system is the free and open source operating system which Debian based and optimized for Raspberry Pi. It provides the basic set of programs and utilities for operating Raspberry Pi. It comes with around 35,000 packages which are pre-compiled software that are bundled in a nice format for hustle free installation on Raspberry Pi. It has good community of developers which runs the discussion forms and provides solutions to many relevant problems. However, Raspbian OS is still under consistent development with an main focus on improving the performance and the stability of as many Debian packages as possible.

V. EXPERIMENTATION AND RESULTS

Figure 5: experimental setup for Node1

As shown in figure 5,experimental setup for node1 consists of mobile robot with central server, GPS module, camera and other sensors. All sensors are successfully interfaced with microcontroller and the microcontroller is interfaced with the raspberry pi.GPS and camera is also connected to raspberry pi. Test results shows that the robot can be controlled remotely using wireless transmission of PC commands to R-Pi. R-Pi forwards the commands to microcontroller and mico controller gives signals to motor driver in order to drive the Robot.GPS module providesthe co-ordinates for the location of the robot

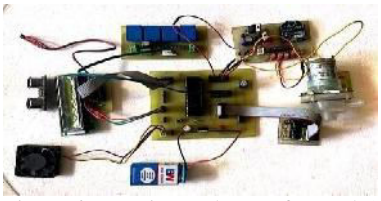


Figure 6: experimental setup for Node2

As shown in above figure, node2 consists of motion detector, temperature sensor, humidity senor, cooling fan, water pump, etc. connected to the microcontroller board.

The sensors gives input to the controller and according to that microcontroller controls the devices in auto mode and also sends the value of sensors to R-Pi and R-Pi forwards it to user's smart device using internet. Test results shows that when temperature level increases above preset threshold level then cooling fan is started automatically in auto mode.

The water pump also gets turned ON if moisture level goes below fixed threshold value. In manual mode, microcontroller receives the controlling signals from R-Pi through ZigBee and accordingly takes thecontrolaction.

Figure 7: experimental setup for Node3

As shown in above figure, node3 consists of a moisture sensor connected to HT12E. Moisture sensor transmits the data using HT12E Encoder IC and a RF transmitter to the Node2 where it is processed by microcontroller and accordingly water pump is switched ON/OFF.

VI. CONCLUSION

The sensors and microcontrollers of all three Nodes are successfully interfaced with raspberry pi and wireless communication is achieved between various Nodes.

All observations and experimental tests proves that project is a complete solution to field activities, irrigation problems, and storage problems using remote controlled robot, smart irrigation system and a smart warehouse management system respectively. Implementation of such a system in the field can definitely help to improve the yield of the crops and overall production.

ACKNOWLEDGMENT

I am sincerely thankful to all my teachers for their guidance for my seminar. Without their help it was tough job for me to accomplish this task. I am especially very thankful to my guide **Dr. R.S. Kawitkar** for his consistent guidance, encouragement and motivation throughout the period of this work. I also want to thank our Head of the Department (E&TC) **Dr. M. B. Mali** for providing me all necessary facilities.

REFERENCES

[1] S. R. Nandurkar, V. R. Thool, R. C. Thool, "Design and Development of Precision Agriculture System Using Wireless Sensor Network", IEEE International Conference on Automation, Control, Energy and Systems (ACES), 2014

[2] JoaquínGutiérrez, Juan Francisco Villa-Medina, Alejandra Nieto- Garibay, and Miguel Ángel Porta-Gándara, "Automated Irrigation System Using a Wireless Sensor Network and GPRS Module",IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, 0018-9456,2013

[3] Dr. V .Vidya Devi,G. Meena Kumari, "Real- Time Automation and Monitoring System for Modernized Agriculture" ,International Journal of Review and Research in Applied Sciences and Engineering (IJRRASE) Vol3 No.1. PP 7-12, 2013

[4] Y. Kim, R. Evans and W. Iversen, "Remote Sensing and Control of an Irrigation System Using a Distributed Wireless Sensor Network", IEEE Transactions on Instrumentation and Measurement, pp. 1379– 1387, 2008.

[5] Q. Wang, A. Terzis and A. Szalay, "A Novel Soil Measuring Wireless Sensor Network", IEEE Transactions on Instrumentation and Measurement, pp. 412–415, 2010

[6] Yoo, S.; Kim, J.; Kim, T.; Ahn, S.; Sung, J.; Kim, D. A2S: Automated agriculture system based on WSN. In ISCE 2007. IEEE International Symposium on Consumer Electronics, 2007,
Irving,TX, USA, 2007

[7] Arampatzis, T.; Lygeros, J.; Manesis, S. A survey of applications of wireless sensors and Wireless Sensor Networks. In 2005 IEEE International Symposium on Intelligent Control & 13th Mediterranean Conference on Control and Automation. Limassol,
Cyprus, 2005, 1-2, 719-724

[8] Orazio Mirabella and Michele Brischetto, 2011. "A Hybrid Wired/Wireless Networking Infrastructure for Greenhouse Management", IEEE transactions on instrumentation and measurement, vol. 60, no. 2, pp 398-407.

[9] N. Kotamaki and S. Thessler and J. Koskiaho and A. O. Hannukkala and H. Huitu and T. Huttula and J. Havento and M. Jarvenpaa(2009). "Wireless in-situ sensor network for agriculture and water monitoring on a river basin scale in Southern Finland: evaluation from a data users perspective". Sensors 4, 9: 2862-2883. doi:10.3390/s90402862 2009.

[10] Liu, H.; Meng, Z.; Cui, S. A wireless sensor network prototype for environmental monitoring in greenhouses. International Conference on Wireless Communications, Networking and Mobile Computing (WiCom 2007), Shangai, China; 21-25 September 2007.

[11] Baker, N. ZigBee and bluetooth - Strengths and weaknesses for industrial applications. Comput. Control. Eng. 2005, 16, 20-25.

[12] IEEE, Wireless medium access control (MAC) and physical layer (PHY) specifications for lowrate wireless personal area networks (LR-WPANs). In The Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2003.

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

Medical Internet Of Things And Bigdata In Healthcare

T.Ramchandarra,G.Naresh

ABSTRACT

A number of technologies can reduce overall costs for the prevention or management of chronic illnesses. These include devices that constantly monitor health indicators, devices that auto-administer therapies, or devices that track real-time health data when a patient self-administers a therapy. Because they have increased access to high-speed Internet and smartphones, many patients have started to use mobile applications (apps) to manage various health needs. These devices and mobile apps are now increasingly used and integrated with telemedicine and telehealth via the medical Internet of Things (mIoT). This paper reviews mIoT and big data in healthcare fields. mIoT is a critical piece of the digital transformation of healthcare, as it allows new business models to emerge and enables changes in work processes, productivity improvements, cost containment and enhanced customer experiences.

I. Introduction

The Internet of Things (IoT) is a network of physical devices and other items, embedded with electronics, software, sensors, and network connectivity, which enables these objects to collect and exchange data [1]. Its impact on medicine will be perhaps the most important, and personal, effect. By 2020, 40% of IoT-related technology will be health-related, more than any other category, making up a $117 billion market [2]. The convergence of medicine and information technologies, such as medical informatics, will transform healthcare as we know it, curbing costs, reducing inefficiencies, and saving lives.

Figure 1 illustrates how this revolution in medicine will look in a typical IoT hospital, in practice. A patient with diabetes will have an ID card that, when

scanned, links to a secure cloud which stores their electronic health record vitals and lab results, medical and prescription histories. Physicians and nurses can easily access this record on a tablet or desktop computer.

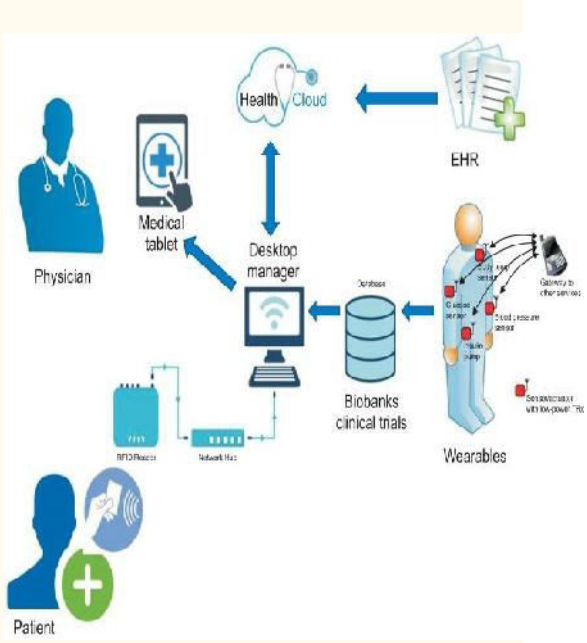


Figure 1
An illustration of how this revolution in medicine will look in a typical Internet of Things (IoT) hospital, in practice.

It sounds pretty basic, but the adoption of Electronic Health Records (EHRs) is a game changer. In less than a decade, an ink-and-paper system of managing records that goes back thousands of years will be digitized and replaced [3]. The advantages The advantages are obvious and many.,

ISBN 978-93-88808-61-3

.One of the major challenges to implementing the IoT has to do with communication; although many devices now have sensors to collect data, they often talk with the server in their own language. Manufacturers each have their own proprietary protocols, which means sensors by different makers can't necessarily speak with each other. This fragmented software environment, coupled with privacy concerns and the bureaucratic tendency to hoard all collected information, frequently maroons valuable info on data islands, undermining the whole idea of the IoT.

Precision medicine, as it's called, is a term that will be frequently heard in coming years [4]. It begins with genomics and goes through the rest of the *omics* platforms, providing multiscale data for analysis and interpretation [5]. In 2015, Intel and the Oregon Health and Science University launched a joint project, the Collaborative Cancer Cloud: a high-performance analytics platform that collects and securely stores private medical data that can be used for cancer research. Though the platform began with cancer, Intel intends to open up the federated cloud network to other institutions, including ones working on cures to diseases like Parkinson's.

Engineering simulation solutions are making medicine participatory, personalized, predictive and preventive (P4 medicine) via the medical Internet of Things (mIoT) [6].

II. IoT - The Future of Pharma?

Pharma companies long ago realized that just selling traditional medicines will not produce growth nor even sustain competitiveness. This fundamental change, known as moving 'beyond the pill', typically arises from one or two realizations: (1) medicines alone are often not enough to achieve optimal clinical outcomes for patients, and (2) as pharmaceutical pipelines dry up, 'beyond-the-pill' businesses can be valuable new sources of revenues. This has created growing interest in methods of utilizing the new technologies and business processes for development and patient care, leading to Pharma IoT.

The Pharma IoT concept involves digitalization of medical products and related care processes using smart connected medical devices and IT services (web, mobile, apps, etc.) during drug development, clinical trials and patient care. The outcomes of Pharma IoT in development and clinical trials can employ combinations of advanced technologies and services to create totally new kinds of disease treatment possibilities (e.g., Treatment 2.0).

In patient care, Pharma IoT will enable patients and healthcare professionals to use medicines with advanced sensor hardware, and craft personalized care services and processes (Product 2.0). Good examples of the Pharma IoT solutions are the connected sensor wearables for Parkinson's disease and multiple sclerosis patients, which provide medication management, improving the patient outcomes and the quality of life [7].

In addition, existing medical device products such as inhalers and insulin pens can be added to the sensor and connectivity technologies to collect data for further care analytics, and even personalized therapy [8]. All this will substantially improve personal medication and care processes, because patient care data provides new sources of innovation and competitiveness.

The transformation also involves some challenges: at the same time, pharma companies need to take into account the forthcoming European Union (EU) data protection and privacy legislation, which will give patients control of their care data

[9]. For example, patients will be allowed to transfer their care and health data across multiple service providers, leading to the emergence of totally new kinds of service platforms and business models, e.g., data brokers [10].

III. Devices and Mobile Apps for Healthcare

We are heading into the age of information, where knowledge and data will be key. We are also entering the age of the customer, in which more than ever the customer is going to determine what they want. *myTomorrows* is one example of the changing look of business models, in this case, directly connecting customers and pharma [11].

In this new age, devices and apps will be used to create a "health selfie". For example:

- *The Myo*, originally a motion controller for games, is now being used in orthopedics for patients who need to exercise after a fracture. With the aid of the Myo, patients can monitor their progress and doctors can measure the angle of movement.
- *The Zio Patch* measures heart rate and electrocardiogram (ECG) and is the US Food and Drug Administration approved [12].

Where is pharma in all this turmoil? Interestingly, there are signs that pharma is reaching out from its traditional medicine-centric approach.

- Glaxo recently announced that it is investing in electroceuticals, bioelectrical drugs that work by micro-stimulation of nerves [13].
- J&J has teamed up with Google to develop robotic surgery. In addition, they are collaborating with Philips on wearable devices such as blood pressure monitors [14].
- Novartis is working with Google (again) on sensor technologies, such as the smart lens, and a wearable device to measure blood glucose levels [15].

Sensors can provide a lot of information to support pharma development, but it is particularly important to recruit the right patients for the right clinical trials. Body sensors, once gadgets that were mainly used by athletes and runners, are now rapidly entering the general market, and consumers and pharma will soon have access to a wealth of information including not only pulse, blood pressure, ECG and respiratory rate, but also more advanced data, such as inflammation, sleep patterns, etc.

A number of mobile apps which support device handling have emerged, including myDario and SleepBot among others [16,17]. The Hacking Medicine Institute recently announced RANKED Health, a program to critically evaluate and rank health-focused applications and connected devices [18].

It has been predicted that in the near future we will look at our phone or smart watch to check health outcomes more often than we do now to check our mail or WhatsApp. A typical situation might involve an elderly person, recovering from a medical condition at home, linked to a combination of several connected services streaming data towards different parties, such as family members, tele-carer and physicians (Figure 2).

ISBN 978-93-88808-61-3

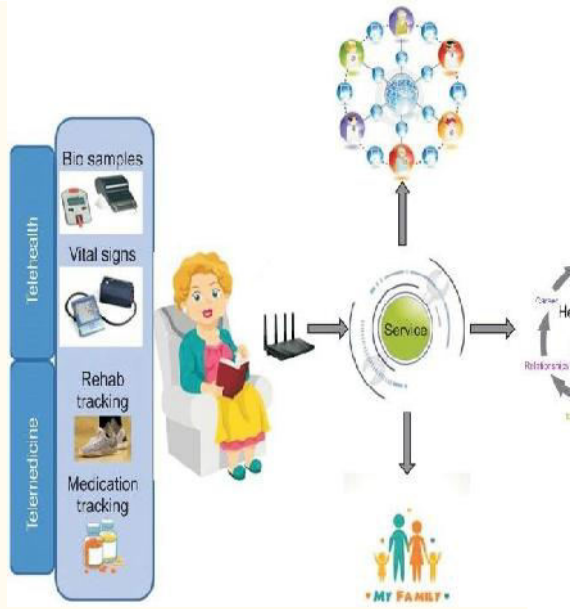


Figure 2
A typical situation involved an elderly person, recovering from a medical condition at home, linked to a combination of several connected services streaming data towards different parties, such as family members, tele-carer and physicians.

Recently it was announced that Medtronic will be partnering with a digital health app company named Canary Health to be a re-seller of its digital chronic disease management programs, including its CDC-recognized Diabetes Prevention Program, which is aimed at changing behaviors in prediabetic people. But the partnership goes beyond just reselling Canary Health's digital tools. In fact, both Canary Health and Medtronic plan to develop solutions that "leverage Medtronic's devices, services and infrastructure as well as Canary Health's suite of behavior-change programs, design expertise, and deep user engagement experience," according to a Canary Health news release [19].

One reason that Medtronic must have been attracted to Canary Health is that the company's digital tools are reimbursable. As digital health programs mature, payers are looking at innovative, yet proven, ways to reduce their cost burden for chronic diseases like diabetes.

According to the Centers for Disease Control and Prevention (CDC), people with prediabetes who take part in a structured lifestyle change program—like the one Canary Health has developed, or programs championed by Omada Health and Noom Health, among others—"can cut their risk of developing type 2 diabetes by 58% (71% for people over 60 years old)" [20]. The CDC adds that "this finding was the result of a program helping people lose 5% to 7% of their body weight through healthier eating and 150 minutes of physical activity a week" [21].

Given that diabetes is an expensive, chronic disease, hospitals, doctors, patients, and payers are equally keen to tame this epidemic. In other words, the move is helping to transform companies from simply providing care to the sick to actually delivering healthcare.

IV. Data

The driver behind all these wearable sensors is the data that is generated, and various parties are trying to bundle the data streams and obtain control. Microsoft developed the Health Vault, an e-health safe, acting as an EMR. In Holland the Radboud University Medical Center collaborated with Philips and Salesforce on *HereIsMyData*, a database where patients can store their health data and determine who can access them [22]. The role of Salesforce is interesting. The Salesforce platform powers Veeva, the customer relationship management (CRM) now widely used in pharma. This positions Salesforce to be able to bridge the gap between patient's medical data and pharma.

"Big data" is a phrase that has been used pervasively by the media and the lay public in the last several years. While

ISBN 978-93-88808-61-3

many definitions have been proposed, the common denominator seems to include the "three V's"—Volume (vast amounts of data), Variety (significant heterogeneity in the type of data available in the set), and Velocity (speed at which a data scientist or user can access and analyze the data) [23].

Defined as such, healthcare has become one of the key emerging users of big data. For example, Fitbit and Apple's ResearchKit can provide researchers access to vast stores of biometric data on users, which can then be used to test hypotheses on nutrition, fitness, disease progression, treatment success, and the like.

Most complex high dimensional data sets include imaging (photos, X-rays, MRIs, and slides), wave analysis such as EEG and ECG, audio files with associated transcripts, free text notes with natural language processing (NLP) outputs, and mappings between structured concepts such as lab tests and the Logical Observation Identifiers Names and Codes (LOINC) codes or the International Classification of Diseases-9 (ICD9) and ICD10 codes. Among the things that the data analysis should provide is the means to continuously update the annotations based on acquired knowledge, while keeping the location of the data in place.

The Centers for Medicare & Medicaid Services (CMS) have vast stores of billing data that can be mined to promote *high value care*; the same is true of private health insurers. And hospitals have attempted to reduce re-admission rates by targeting patients where predictive artificial intelligence (AI) algorithms indicate people who may be at highest risk based on an analysis of available data collected from existing patient records

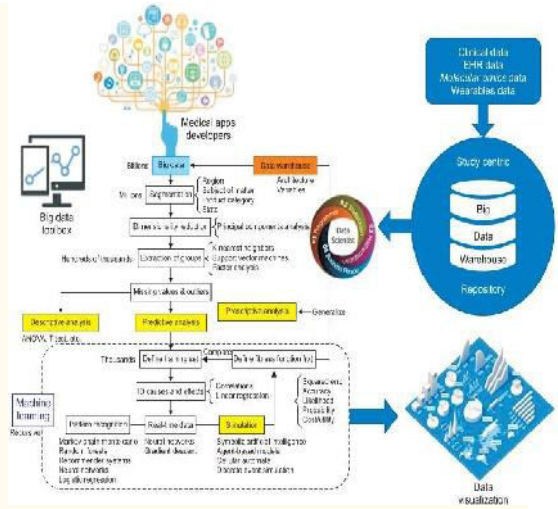


Figure 3
The Centers for Medicare & Medicaid Services (CMS) data system.

Underlying these and many other potential uses, however, are a series of technology, legal and ethical challenges relating to, among other things, privacy, discrimination, intellectual property, tort, and informed consent, as well as research and clinical ethics [24].

V. Challenges for mIoT

Leading IoT platforms must provide simple, powerful application access to IoT devices and data to help designers rapidly compose analytics applications, visualization dashboards and mIoT apps. The following are 5 key capabilities that leading platforms must enable:

(1) Simple connectivity: A good IoT platform makes it easy to connect devices and perform device management functions, scaled through cloud-based services, and to apply analytics to gain insight and achieve organizational transformation.

(2) Easy device management: A thoughtful approach to device management enables improved asset availability, increased

throughput, minimized unplanned outages and reduced maintenance costs.

(3) Information ingestion: Intelligently transform and store IoT data. APIs bridge the divide between the data and the cloud, making it easy to pull in the data that's needed. Data is ingested from diverse data sources and platforms, then the essential values are extracted using rich analytics.

(4) Informative analytics: Gain insight from huge volumes of IoT data to make better decisions and optimize operations. Apply real-time analytics to monitor current conditions and respond accordingly. Leverage cognitive analytics with both structured and unstructured data to understand situations, reason through options, and learn as conditions change. An intuitive dashboard makes it all easy to understand.

(5) Reduced risk: Act on notifications and isolate incidents generated anywhere in the company environment from a single console.

VI. Challenges for Big Data in Healthcare

The challenges fall into two main categories: fiscal/policy and technology.

Fiscal and policy issues: In a fee-for-service environment, the only way that healthcare practitioners get paid is to have face-to-face encounters with patients. This creates heavy bias against promoting technologies that streamline non-face-to-face interactions. However, as we move away from that model and more towards value-based care, where global risk-based payments are made to delivery organizations (hospitals, patient centered medical homes, accountable care organizations, etc.), then there is more incentive to use new technologies that reduce unnecessary in-office encounters. In such an environment, face-to-face encounters are actually a cost center, not a profit center, and positive health outcomes of populations are rewarded.

Technology issues: The biggest technical barrier to achieving this vision is the state of health data. Created by legacy EHR systems, health data is largely fragmented into institution-centered silos. Sometimes those silos are large, but they are still silos. Exchanging individual records between silos, using increasingly standardized vocabularies (code sets) and message formats (ADT messages, C-CDAs, even FHIR objects), is where much current effort is being directed. But that does not solve the problem of data fragmentation. More and more people in the health information exchange arena are seeing that the next generation of health technology is around aggregating data, not simply exchanging copies of individual records (the traditional query-response approach). Only by collecting the data from many different sources, normalizing that data into a consistent structure, resolving the data around unique patient identifiers as well as unique provider identifiers—only then can the data become truly useful [25].

Aggregated data has two additional advantages. (1) It solves the interoperability problem. Systems and institutions no longer need to build data bridges, and translate how the data is structured between two proprietary systems; everyone instead simply connects to a central standard API "plug." If built right, the aggregated data can be the basis for very effective AI technology. Such technology is very fast (consider Google suggestions as-you-type in a search bar, retrieving suggestions from billions of record options). (2) It is also sufficiently flexible to allow machine learning, and AI will be able to function in a real-time fashion.

VII. New Generation of Digital Health Advisors

Once a data store has been built from many different sources—EHR data, payer data, device and IoT data, patient survey responses, consumer health data—and has been integrated into a unified data structure, then AI can yield meaningful insights. AI, after all, is about pattern recognition, comparing a particular pattern of data around a given individual with similar (not necessarily identical) patterns found elsewhere, and making predictive recommendations based on what happened in those other situations. This is very much what clinicians do when exercising "clinical judgement"—identifying a pattern, taking into account medical problems, medications, labs values, personal and family history, and comparing it to similar patterns from the clinician's experience.

A new generation of "Health Coaches", Tele-Carers or Digital Health Advisors can be trained to make these AI-derived recommendations useful [26]. They need to be easy-to-use, consumer-orientated persons who can connect to the aggregated data store and the AI analytics engines that sit on top of that. They can empower consumers/patients, and reduce the demand burden on clinicians. Will they replace clinicians? No, of course not. But they will help filter the demand to those who truly need to be seen, while empowering patients with real-time, believable and personalized guidance for the more common things in day-to-day life [27].

So what stands in the way of Digital Health Advisors? Policy (how we pay for healthcare) needs to encourage self-care and facilitate healthy behaviors, rather than encourage inoffice doctor visits. And, simultaneously, health data needs to become reorganized in order to empower AI and drive the emergence of new apps and related technologies. It will be a while before we get there, but we can see the path to that new generation of healthcare technology.

Concliousn

The mIoT is revamping healthcare services, as people have started using IoT to manage their health requirements. For example, people can use IoT devices to remind them about appointments, changes in blood pressure, calories burnt and much more. One of the best parts of the IoTs in the healthcare industry is the remote health monitoring system, where patients can be monitored and advised from anywhere. Real-time location services are another major approach IoT offers. By using the service, doctors can easily track device locations, which directly reduces excess time spent. Smartphone usage is increasing rapidly, and people have started using mobile apps for almost everything. When it comes to the healthcare industry, mobile apps can improve communications between patients and doctors over a secured connection.

The primary duty of Digital Health Advisors and the clinicians will be to work collaboratively when the organization is shifting towards IoT-enabled infrastructure. Proper training and feedback are mandatory for better deployment. The traditional method of recording a patient's details, i.e., a pad of paper hanging on the patient's bed, is not going to work anymore, since such records are only accessible to a limited few, and can be lost or scrambled. This is an application where on-field mobile/tablet technology might work, since they offer hassle-free record management on the applications in the device. Health data information will be available in just a tap when information is recorded electronically, once security and privacy issues are met.

:

References

1. Zanella A, Bui N, Castellani A, Vangelista L, Zorzi M. Internet of things for smart cities. IEEE Internet Things J. 2014;1(1):22–32. [Google Scholar]
2. Bauer H, Patel M, Veira J. The Internet of Things: sizing up the opportunity [Internet] New York (NY): McKinsey & Company; c2016. [cited at 2016 Jul 1]. Available from: http://www.mckinsey.com/industries/high-tech/our-insights/the-internet-of-things-sizing-up-the-opportunity. [Google Scholar]
3. Kruse CS, Kothman K, Anerobi K, Abanaka L. Adoption factors of the electronic health record: a systematic review. JMIR Med Inform. 2016;4(2):e19. [PMC free article] [PubMed] [Google Scholar]
4. Scheen AJ. Precision medicine: the future in diabetes care? Diabetes Res Clin Pract. 2016;117:12–21. [PubMed] [Google Scholar]
5. van Leeuwen N, Swen JJ, Guchelaar HJ, 't Hart LM. The role of pharmacogenetics in drug disposition and response of oral glucose-lowering drugs. Clin Pharmacokinet. 2013;52(10):833–854. [PubMed] [Google Scholar]
6. Flores M, Glusman G, Brogaard K, Price ND, Hood L. P4 medicine: how systems medicine will transform the healthcare sector and society. Per Med. 2013;10(6):565–576. [PMC free article] [PubMed] [Google Scholar]
7. van Uem JM, Maier KS, Hucker S, Scheck O, Hobert MA, Santos AT, et al. Twelve-week sensor assessment in Parkinson's disease: impact on quality of life. Mov Disord. 2016 May 31; doi: 10.1002/mds.26676. [Epub] [PubMed] [CrossRef] [Google Scholar]
8. Dzubur E, Li M, Kawabata K, Sun Y, McConnell R, Intille S, et al. Design of a smartphone application to monitor stress, asthma symptoms, and asthma inhaler use. Ann Allergy Asthma Immunol. 2015;114(4):341–342.e2. [PMC free article] [PubMed] [Google Scholar]
9. European Commission. Clinical trials - regulation EU No 536/2014 [Internet] Brussels: European Commission; c2016. [cited at 2016 Jul 1]. Available from: http://ec.europa.eu/health/human-use/clinical-trials/regulation/index_en.htm. [Google Scholar]
10. INNT Foundation [Internet] Amsterdam: INNT Foundation; c2016. [cited at 2016 Jul 1]. Available from: https://www.innit.foundation. [Google Scholar]
11. MyTomorrows [Internet] Amsterdam: MyTomorrows; c2016. [cited at 2016 Jul 1]. Available from: https://mytomorrows.com. [Google Scholar]
12. Tung CE, Su D, Turakhia MP, Lansberg MG. Diagnostic yield of extended cardiac patch monitoring in patients with stroke or TIA. Front Neurol. 2015;5:266. [PMC free article] [PubMed] [Google Scholar]
13. Famm K, Litt B, Tracey KJ, Boyden ES, Slaoui M. Drug discovery: a jump-start for electroceuticals. Nature. 2013;496(7444):159–161. [PMC free article] [PubMed] [Google Scholar]
14. Cuba-Gyllensten I, Gastelurrutia P, Riistama J, Aarts R, Nunez J, Lupon J, et al. A novel wearable vest for tracking pulmonary congestion in acutely decompensated heart failure. Int J Cardiol. 2014;177(1):199–201. [PubMed] [Google Scholar]
15. Senior M. Novartis signs up for Google smart lens. Nat Biotechnol. 2014;32(9):856. [PubMed] [Google Scholar]
16. MyDario.com [Internet] Burlington (MA): MyDario.com; c2016. [cited at 2016 Jul 1]. Available from: http://mydario.com/ [Google Scholar]

17. SleepBot [Internet] New York (NY): SleepBot; c2013. [cited at 2016 Jul 1]. Available from: https://mysleepbot.com/ [Google Scholar]
18. RANKED Health [Internet] place unknown: publisher unknown; [cited at 2016 Jul 1]. Available from: http://www.rankedhealth.com/about/ [Google Scholar]
19. Canary Health [Internet] Los Angeles (CA): Canary Health Inc.; c2016. [cited at 2016 Jul 1]. Available from: http://www.canaryhealth.com/ [Google Scholar]
20. Sepah SC, Jiang L, Peters AL. Long-term outcomes of a Web-based diabetes prevention program: 2-year results of a single-arm longitudinal study. J Med Internet Res. 2015;17(4):e92. [PMC free article] [PubMed] [Google Scholar]
21. Centers for Disease Control and Prevention. Lifestyle change program details [Internet] Atlanta (GA): Centers for Disease Control and Prevention; c2016. [cited at 2016 Jul 1]. Available from:]
22. Hhttp://www.cdc.gov/diabetes/prevention/lifestyle-program/experience/index.html. [Google ScholarereIsMyData [Internet] Nijmegen: Radboud University Medical Center; c2015. [cited at 2016 Jul 1]. Available from: http://www.hereismydata.com/ [Google Scholar]
23. Auffray C, Balling R, Barroso I, Bencze L, Benson M, Bergeron J, et al. Making sense of big data in health research: towards an EU action plan. Genome Med. 2016;8(1):71. [PMC free article] [PubMed] [Google Scholar]
24. Filkins BL, Kim JY, Roberts B, Armstrong W, Miller MA, Hultner ML, et al. Privacy and security in the era of digital health: what should translational researchers know and do about it? Am J Transl Res. 2016;8(3):1560–1580. [PMC free article] [PubMed] [Google Scholar]
25. Hackl WO. Intelligent re-use of nursing routine data: opportunities and challenges. Stud Health Technol Inform. 2016;225:727–728. [PubMed] [Google Scholar]
26. Foley P, Steinberg D, Levine E, Askew S, Batch BC, Puleo EM, et al. Track: a randomized controlled trial of a digital health obesity treatment intervention for medically vulnerable primary care patients. Contemp Clin Trials. 2016;48:12–20. [PMC free article] [PubMed] [Google Scholar]
27. Giraldo-Rodriguez L, Torres-Castro S, Martinez-Ramirez D, Gutierrez-Robledo LM, Perez-Cuevas R. Tele-care and tele-alarms for the elderly: preliminary experiences in Mexico. Rev Saude Publica. 2013;47(4):711–717. [PubMed] [Google Scholar]

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

Prediction of Cardiac Arrest in Intensive Care Patients through Machine Learning

Rahmath Unissa and Shiva Krishna P

Abstract
Cardiac arrest is a critical health condition characterized by absence of traceable heart rate, patient's loss of consciousness as well as apnea, with inhospital mortality of *80%. Accurate estimation of patients at high risk is crucial to improve not only the survival rate, but also the quality of life as patients who survived from cardiac arrest have severe neurological effects. Existing research has focused on demonstrating static risk scores without taking account patient's physiological condition. In this study, we are implement-ing an integrated model of sequential contrast patterns using Multichannel Hidden Markov Model. These models can capture relations between exposure and control group and offer high specificity results, with an average sensitivity of 78%, and have the ability to identify patients in high risk.

Keywords
Cardiac arrest Prediction MC-HMM Sequential pattern recognition Classification

Introduction

Cardiac arrest is defined as interruption of mechanical activity of heart, which is confirmed by absence of traceable heart rate, patient's loss of consciousness, as well as apnea, according to the Utstein style. Cardiac arrest is defined as inpatient when it occurs in a hospitalized patient who had pulse on admission to the hospital [1]. Common causes of cardiac arrest are ventricular fibrillation (VF), ventricular tachycardia (VT), asystole and electrical activity of the heart without pulses. About 200,000 cases of inpatient cardiac arrest are reported each year in U.S.A. (United States of America) [2]. Cardiac arrest occurs in 1–5 per 1000 hospi-talized patients and *20% survive until their discharged [1, 3, 4]. Generally, patients at high risk of cardiac arrest have co-morbidities, which affect their health outcome and recovery after cardiac arrest [2]. Studies have shown that clinical signs of deterioration, such as hemodynamic insta-bility and respiratory distress, of patients within a period of eight hours prior to cardiac arrest could be used to avoid cardiac arrest in 84% of these [1]. However, the recognition of the causes of cardiac arrest, has been shown to increase the survival rate of patients within an hour of episode by about 29% and by 19% until their discharge [3]. Therefore, early and accurate detection of patients at-risk is critical to improve health outcome and survival rate.

Increasing use of electronic health records (EHR) leads to greater accessibility and availability of medical data. The

E. Akrivos (&) N. Maglaveras I. Chouvarda
Lab of Computing Medical Informatics and Biomedical Imaging Technologies, Medical School, Aristotle University of Thessaloniki, Thessaloniki, Greece e-mail: e_akrivos@icloud.com

E. Akrivos
Department of Internal Medicine, 424 Military Hospital of Thessaloniki, Thessaloniki, Greece

V. Papaioannou
Alexandroupolis University Hospital/Intensive Care Unit, Alexandroupoli, Greece

N. Maglaveras
McCormick School of Engineering & Applied Sciences, Department of Electrical Engineering & Computer Science, Northwestern University, Evaston, IL 60201, USA

Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC II) database was developed from medical data of over 30,000 patients during 2001–2008 from Beth Israel Deaconess Medical Center in Boston. MIMIC II is the most extensive resource of intensive care unit (ICU) medical data and it is available to the public [5, 6].

Recent research used measurements of vital signs, such as blood pressure, respiratory rate, temperature and healthcare professional's opinion to model early warning scores to identify patients at high risk of cardiac arrest [7–9]. However, these researches could not predict the accurate time of cardiac arrest. DYNACARE is a model based on dynamic time series attending to predict the time of cardiac arrest [10].

The present study proposes an approach that discovers sequential contrast patterns from commonly observed measurements, such as blood pressure, respiratory rate and heart rate, transforming the classical time series data to a sequence of patterns for implementation of a classifier for cardiac arrest. Following, the classifier is used to predict the likeli-hood of a sequence of patterns to belong in cardiac arrest class. This method has been used for the prediction of Sepsis [11], but to our knowledge it is now applied to cardiac arrest prediction for the first time.

Materials and Methods

The study was conducted with data from MIMIC-II for adult patients (age 18+ on ICU admission) aged up to 90 years who were hospitalized in the Cardiological ICU and expe-rienced a recorded cardiac arrest episode according to ICD-9 (International Classification of Diseases) 427.5 for cardiac arrest. The study focused on different types of variables, such as demographic data, vitals signs, medication and laboratory measurements. Patient data was discretized in 2-hour bins. An additional requirement for each patient was to have at least 36 measurements (3 days of hospitalization) to ensure sufficient data points. There were 698 patients with a cardiac arrest diagnosis from 27,542 of MIMIC-II data-base, from which only 162 met the minimum data criteria. Patients who have been diagnosed with highrisk heart dis-eases for cardiac arrest and have not occured an event of cardiac arrest, were selected as control group, with diseases such as coronary heart disease, myocardial infarction, major heart disease, valvular heart disease, congenital heart disease and heart rhythm abnormalities such as Brugada syndrome and long QT [12–22]. The selected ICD-9 codes for these diseases was 414.01, 410.90, 429.3, 424.0, 424.1, 746.0– 746.9, 746.89 and 426.82. Similar data criteria to patients with cardiac arrest were also used in control population, with a final number of control population 5,278 patients.

Data Preparation and Preprocessing

The first step was extraction of data from Mimic Database in flat files. Quality inspection revealed a number of missing values in different data fields. Missing data was processed using the Multiple Imputation method and predictive mean matching (PMM) algorithm [23]. In order to increase similarity of considered cases, medication, demographic data and laboratory measurements were used as coefficients for PMM to predict missing values of heart rate, systolic blood pressure, diastolic blood pressure, respiratory rate, PO_2 and PCO_2.

Following, the quantization of measurements and map-ping to specific states was necessary for sequential pattern analysis, since pattern discovery methods are more effective on symbolic data types. Frequent sequence patterns methods [24–26] is used to identify patterns and frequency support in sequences between the two classes of sequence data.

Mining Sequential Contrast Patterns

Emerging patterns (EPs) are described as patterns that satisfied specific user-defined frequency rules for different classes of data. This means that in a categorized data in two categories, positive (cardiac arrest group) and negative (control group), the patterns must have a high frequency support in the positive category and a low frequency support in the negative category. Since EPs have these characteris-tics, they are considered to be distinct patterns and have the ability to distinguish the contrast between the two categories (also known as growth rate of EP). Therefore, the strength of EPs is expressed by the ratio of frequency in both classes.

Extending the above description, a sequence pattern S_p can be characterized as sequential contrast pattern if satisfy the conditions (a) and (b) depicted below in Eqs. (1) and (2)

(a) Positive support:

$$counts_{S_p}(D^{+}; g) \qquad \text{a}$$

(b) Negative support:

$$counts_{S_p}(D^{-}; g) \qquad \text{b}$$

where D^{+} , D^{-} two different datasets with labels, such as positive sequences and negative sequences, respectively, g is the gap-constraint, $counts_{S_p}(D; g)$ the frequency support of a

Table 1 Contrast patterns for heart rate sequences

Heart Rate patterns	Pattern-id
Tachyc < N1hr	HR1
Tachyc < Nhr	HR2
Tachyc < N1hr < N1hr	HR3
Tachyc < Tachyc	HR4

Table 2 Contrast patterns for systolic blood pressure sequences

Systolic BP patterns	Pattern-id
Nbpsys < HypotensS	SB1
HypotensS < Nbpsys	SB2
Nbpsys < Nbpsys < HypotensS	SB3

Table 3 Contrast patterns for respiratory rate sequences

Respiratory Rate patterns	Pattern-id
Bradypnoea-FP-FP	RR1
FP-FP-FP	RR2
FP-FP-Bradypnoea	RR3

sequence pattern S_p, a and b thresholds for frequency support in two datasets. Thus, discovered patterns lead to mining sequential contrast patterns, given the above characteristics, which must satisfy (a) and (b) condition [11].

In the present study, using the above description, after descretization of variables based on normal value's cut offs, resulted in contrast patterns for three variables, where a = 0.7, b = 0.5 and g = 2. Variables with contrast patterns were heart rate, systolic blood pressure and respiratory rate. Tables 1, 2 and 3 show the contrast patterns and their unique identification name with which they were replaced.

A sliding window with length equivalent to the longest pattern (length = 3) was used to transform the discrete sequences of data to sequences of contrast patterns, with purpose to use these as input data to HMM, instead of ordinary time series sequences. Table 4 shows the above transformation from a patient's sequences.

Multichannel Hidden Markov Model

Multi Channel Hidden Markov Models (MC-HMM) are an extension of the conventional form of Hidden Markov Models (HMMs) for multiple variable or channel data sequences. MC-HMM has been used on applications such as speech recognition, activity recognition, anomalous trading activities, medical events, disease interactions and fault diagnosis [27–30]. In the present study, MC-HMM was used to model interactions between multiple clinical measurements, which are represented as sequential contrast patterns. According to the theory of MC-HMM, each discrete state for each channel is individually transformed into a three-state mode, based on the markov property. Therefore, it appears that the probability of transition and emission for each state can be mapped as a mutation of the three unique states that correspond to each channel. Two MC-HMM's constructed for the two classes of data. The first MC-HMM was trained by expectation maxi-mization (EM) algorithm for patients who belong in cardiac arrest class, while the second one was trained for patients of control group.

Results

For prediction of cardiac arrest, the 8-fold cross validation method performed. For each dataset, cardiac arrest and control dataset respectively, 7 folds randomly selected used as training datasets for each model respectively and 1 fold for test set. Thus, each model was trained to find the sequences that belong to their class. Test set from cardiac arrest patient's data was containing only the sequences from observational window before the onset of cardiac arrest, for the classification purposes. Test sets from the two datasets were merged and likelihood for each patient's sequence computed for the two models. If the likelihood of the sequence patterns of the cardiac arrest patient's model was greater than the control pattern then the patient was considered to belong in class with patients at higher risk of

Table 4 Example of transformation from discrete patient sequence to sequence of contrast pattern

Discrete sequences	Variable	Contrast pattern ID
N1hr-N1hr-Tachyc-N1hr-Nhr-N1hr-….-Nhr-Tachyc-N1hr-N1hr-N1hr-Nhr-Nhr-N1hr	HR	HR1-HR3-HR2-HR3-X-X-X-X
HypotensS-Nbpsys-Nbpsys-Nbpsys-…-Nbpsys-HypotensS-Nbpsys-Nbpsys-Nbpsys-Nbpsys	SysBP	SB2-SB3-SB1-SB2-X-X-X-X
Bradypnoea-FP-RRnorm-FP-RRnorm-…-RRnorm-RRnorm-FP-RRnorm-FP-Bradypnoea-FP-RRnorm	RR	RR1-RR3-X-X-X-X-X-X

Table 5 Statistical results for 8 fold cross validation prediction

Sensitivity (mean ± SD)	Specifity (mean ± SD)
0.78 (±0.04)	0.43 (±0.02)

cardiac arrest and were categorized respectively. Table 5 shows statistical results from prediction.

Discussion

The study's results have evidence that integrating MC-HMM models with sequential contrast patterns as input data can perform well to predict cardiac arrest. A limitation in selection criteria of cardiac arrest group, which led to the reduced sample, was that patients with respiratory cardiac arrest were excluded. By discovering patterns, based on the contrast of their frequencies on data of two populations, intervention and control respectively, it is possible to interpret the difference between the two data populations. In particular, the use of a and b thresholds to calculate the growth rate of a pattern is equivalent to the odds ratio, which is used in medical research to find relationships between an exposure and an outcome. However, false positive rate is high. This issue was the result of the restricted design of population groups with ICD-9 code. In the present study, cardiac arrest diagnosis was one of the criteria for patient selection, while VT and VF diagnosis were criteria for control group. This issue poses the problem of semantically defining and selecting the correct cases within a rich database. Furthermore, the significance of a contrast pattern is determined by the growth rate, which if it is too high it creates few patterns, and if it is too low it creates patterns without significance. Therefore, creating an algorithm for optimal selection of the threshold value for pattern development is necessary in order to find important contrast patterns.

Conclusion

In this study, an attempt was made to model cardiac arrest by using an integrated framework, which was previously successfully tested to predict septic shock [11]. However, while the results are promising, it became obvious that the complexity of cardiac arrest mechanism poses many difficulties in modeling. Thus, the present study demonstrates the importance of using sequential contrast patterns to capture relations between groups.

Conflict of Interest The authors declared no potential conflicts of interest with respect to the authorship and/or publication of this article.

References

1. Sandroni C, Nolan J, Cavallaro F, Antonelli M (2007) In-hospital cardiac arrest: incidence, prognosis and possible measures to improve survival. Intensive Care Med 33(2):237–245
2. Graham R, McCoy, MA, Schultz AM (2015) Committee on the treatment of cardiac arrest: current status and future directions, Board on Health Sciences Policy, Institute of Medicine Strategies to improve Cardiac Arrest Survival: A Time to Act. Washington (DC). National Academies Press (US), 29 Sept 2015
3. Bergum D, Haugen BO, Nordseth T, Mjølstad OC, Skogvoll E (2015) Recognizing the causes of in-hospital cardiac arrest-A survival benefit. Resuscitation. 97:91–96
4. Nolan JP, Soar J, Smith GB, Gwinnutt C, Parrott F, Power S et al (2014) Incidence and outcome of in-hospital cardiac arrest in the United Kingdom NationalCardiac Arrest Audit. Resuscitation. 85 (8):987–992
5. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG et al (2000) Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. Circulation 101(23):E215–E220
6. Saeed M, Villarroel M, Reisner AT, Clifford G, Lehman LW, Moody G et al (2011) Multiparameter intelligent monitoring in intensive care ii: a public-access intensive care unit database. Crit Care Med 39(5):952–960
7. Smith AF, Wood J (1998) Can some in-hospital cardiorespiratory arrests be prevented? A prospective survey. resuscitation. 37 (3):133–137
8. Hodgetts TJ, Kenward G, Vlachonikolis IG, Payne S, Castle N (2002) The identification of risk factors for cardiac arrest and formulation of activation criteria to alert a medical emergency team. Resuscitation. 54(2):125–131
9. McBride J, Knight D, Piper J, Smith GB (2005) Long-term effect of introducing an early warning score on respiratory rate charting on general wards. Resuscitation. 65(1):41–44
10. Ho JC, Park Y, Carvalho CM, Ghosh J (2013) DYNACARE: dynamic cardiac arrest risk estimation. J Mach Learn Res 31:333–341

11. Ghosh S, Li J, Cao L, Ramamohanarao K (2017) Septic shock prediction for ICU patients via coupled HMM walking on sequential contrast patterns. J Biomed Inf 66:19–31
12. Longo DL et al (2015) Cardiovascular collapse, cardiac arrest, and sudden cardiac death. In: Harrison's Principles of Internal Medicine, 19th edn, New York
13. Sudden cardiac arrest. http://www.nhlbi.nih.gov/health/health-topics/topics/scda/
14. Podrid PJ. Overview of sudden cardiac arrest and sudden cardiac death. http://www.uptodate.com/home
15. American Heart Association. Heart attack or sudden cardiac arrest: How are they different? http://www.heart.org/HEARTORG/Conditions/More/MyHeartandStrokeNews/Heart-Attack-or-Sudden-Cardiac-Arrest-How-Are-They-Different_UCM_440804_ Article.jsp-Vi55p36rTIU
16. Neumar RW, Shuster M, Callaway CW, Gent LM, Atkins DL et al (2015) Part 1: executive summary: 2015 American heart associ-ation guidelines update for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation. 132(18 Suppl 2): S315-167
17. Arrhythmia. National Heart, Lung, and Blood Institute. http:// www.nhlbi.nih.gov/health/health-topics/topics/arr
18. Fuster V et al (2011) Sudden cardiac death in hurst's the heart, 13th edn. The McGraw-Hill Companies, New York
19. Goldberger AL et al (2013) Sudden cardiac arrest and sudden cardiac death in clinical electrocardiography: a Simplified Approach, 8th edn. Saunders Elsevier, Philadelphia
20. Association. AH. Ejection fraction heart failure measurement. http://www.heart.org/HEARTORG/Conditions/HeartFailure/SymptomsDiagnosisofHeartFailure/Ejection-Fraction-Heart-Failure-Measurement_UCM_306339_Article.jsp-Vi58RH6rTIU
21. Riggin EA. Allscripts EPSi. Mayo Clinic, Rochester, Minn
22. Rohren CH (expert opinion). Mayo Clinic, Rochester, Minn
23. Vink G, Frank LE, Pannekoek J, van Buuren S (2014) Predictive mean matching imputation of semicontinuous variables. Stat Neerl 68(1):61–90
24. Klema J, Novakova L, Karel F, Stepankova O (2008) Sequential data mining: A comparative case study in development of atherosclerosis risk factors. Syst Man Cybern Part C Appl Rev IEEE Trans 38(1):3–15
25. Baralis E, Bruno G, Chiusano S, Domenici VC, Mahoto NA, Petrigni C (2010) Analysis of medical pathways by means of frequent closed sequences. In: Knowledge-based and intelligent information and engineering systems, pp. 418–425
26. Berlingerio M, Bonchi F, Giannotti F, Turini F (2007) Time-annotated sequences for medical data mining. In: Seventh IEEE international conference on data mining workshops. IEEE, pp 133–138
27. Audhkhasi K, Osoba O, Kosko B (2013) Noisy hidden Markov models for speech recognition. In: The 2013 international joint conference on neural networks (IJCNN)
28. Cao L, Ou Y, Yu PS (2012) Coupled behavior analysis with applications. IEEE Trans Knowl Data Eng 24(8):1378–1392
29. Masoudi S, Montazeri N, Shamsollahi MB, Ge D, Beuche A, Pladys P, et al (2013) Early detection of apnea-bradycardia episodes in preterm infants based on coupled hidden Markov model. IEEE international symposium on signal processing and information technology
30. Zhou H, Chen J, Dong G, Wang H, Yuan H (2016) Bearing fault recognition method based on neighbourhood component analysis and coupled hidden Markov model. Mech Syst Signal Process 66– 67:568–581

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

An Efficient and Privacy-Preserving Biometric Identification Scheme in Cloud Computing

K.Ashok, M.Tech
Scholar, CSE
Department,
Malla Reddy College of Engineering,
k.ashok.kanneboina@gmail.com

ABSTRACT Biometric identification has become increasingly popular in recent years. With the develop- ment of cloud computing, database owners are motivated to outsource the large size of biometric data and identification tasks to the cloud to get rid of the expensive storage and computation costs, which however brings potential threats to users' privacy. In this paper, we propose an efficient and privacy-preserving biometric identification outsourcing scheme. Specifically, the biometric data is encrypted and outsourced to the cloud server. To execute a biometric identification, the database owner encrypts the query data and submits it to the cloud. The cloud performs identification operations over the encrypted database and returns the result to the database owner. A thorough security analysis indicates the proposed scheme is secure even if attackers can forge identification requests and collude with the cloud. Compared with previous protocols, experimental results show the proposed scheme achieves a better performance in both preparation and identification procedures.

KEYWORDS: biometric identification; data outsourcing; privacy-preserving; cloud computing

I. INTRODUCTION

BIOMETRIC identification has raised increasingly attention since it provides a promising way to identify users. Compared with traditional authentication methods based on passwords and identification cards, biometric identification is considered to be more reliable and convenient [1]. Additionally, biometric identification has been widely applied in many fields by using biometric traits such as fingerprint [2], iris [3], and facial patterns [4], which can be collected from various sensors [5]–[9].

In a biometric identification system, the database owner such as the FBI who is responsible to manage the national fingerprints database, may desire to outsource the enormous biometric data to the cloud server (e.g., Amazon) to get rid of the expensive storage and computation costs. However, to preserve the privacy of biometric data, the biometric data has to be encrypted before outsourcing. Whenever a FBI's partner (e.g., the police station) wants to authenticate an individual's identity, he turns to the FBI and generates an identification query by using the individual's biometrictraits (e.g., fingerprints, irises, voice patterns, facial patterns etc.). Then, the FBI encrypts the query and submits it to the cloud to find the close match. Thus, the challenging problem is how to design a protocol which enables efficient and privacy-preserving biometric identification in the cloud computing.

A number of privacy-preserving biometric identification solutions [10]–[17] have been proposed. However, most of them mainly concentrate on privacy preservation but ignore the efficiency, such as the schemes based on homomorphic encryption and oblivious transfer in [10], [11] for fingerprint and face image identification respectively. Suffering from performance problems of local devices, these schemes are not efficient once the size of the database is larger than 10 MB. Later, Evans et al. [12] presented a biometric identification scheme by utilizing circuit design and ciphertext packing techniques to achieve efficient identification for a

larger database of up to 1GB. Additionally, Yuan and Yu [13] proposed an efficient privacy-preserving biometric identification scheme. Specifically, they constructed three modules and designed a concrete protocol to achieve the security of fingerprint trait. To improve the efficiency, in their scheme, the database owner outsources identification matching tasks to the cloud. However, Zhu et al. [18] pointed out that Yuan and Yu's protocol can be broken by a collusion attack launched by a malicious user and cloud. Wang et al. [14] proposed the scheme CloudBI-II which used random diagonal matrices to realize biometric identification. However, their work was proven insecure in [15], [16].

In this paper, we propose an efficient and privacy-preserving biometric identification scheme which can resist the collusion attack launched by the users and the cloud. Specifically, our main contributions can be summarized as follows:

- We examine the biometric identification scheme [13] and show its insufficiencies and security weakness under the proposed level-3 attack. Specifically, we demonstrate that the attacker can recover their secret keys by colluding with the cloud, and then decrypt the biometric traits of all users.
- We present a novel efficient and privacy-preserving biometric identification scheme. The detailed security analysis shows that the proposed scheme can achieve a required level of privacy protection. Specifically, our scheme is secure under the biometric identification outsourcing model and can also resist the attack proposed by [18].
- Compared with the existing biometric identification schemes, the performance analysis shows that the proposed scheme provides a lower computational cost in both preparation and identification procedures.

The remainder of this paper is organized as follows: section II presents the models and design goals. In section III, we provide an overview and the security analysis of the previous protocol proposed by Yuan and Yu. In section IV, we present an efficient and privacy-preserving biometric identification scheme. Security analysis is presented in section V, followed by performance evaluation in section VI. In section VII, we give the related work and we show our conclusions in section VIII.

II. MODELS AND DESIGN GOALS

This section introduces the system model, attack model, design goals and the notations used in the following sections.

A. SYSTEM MODEL

As shown in Fig.1, three types of entities are involved in the system including the database owner, users and the cloud. The database owner holds a large size of biometric data (i.e., fingerprints, irises, voice, and facial patterns etc.), which is encrypted and transmitted to the cloud for storage. When a user wants to identify himself/herself, a query request is be sent to the database owner. After receiving the request, the database owner generates a ciphertext for the biometric trait and then transmits the ciphertext to the cloud for identification. The cloud server figures out the best match for the encrypted query and returns the related index to the database owner. Finally, the database owner computes the similarity between the query data and the biometric data associated with the index, and returns the query result to the user.

In our scheme, we assume that the biometric data has been processed such that its representation can be used to execute biometric match. Without loss of generality, similar to[17],[18], we target fingerprintsanduse FingerCodes [19] to represent the fingerprints. More specifically, a FingerCode consists of n elements and each element is a l-bit integer (typically $n = 640$ and $l = 8$). Given two FingerCodes $x = [x_1, x_2, .,. .x_n]$ and $y = [y_1, y_2, .,. .y_n]$, if their Euclidean distance is below a threshold s, they are usually considered as a good match, which means the two fingerprints are considered from the same person.

B. ATTACK MODEL

First of all, the cloud server is considered to be "honest but curious" as described in [13]–[15], [17]. The cloud strictly follows the designed protocol, but makes efforts to reveal privacy from both the database owner and the user. We assume that an attacker can observe all the data stored in the cloud including the encrypted biometric database, encrypted queries and matching results. Moreover, the attacker can act as a user to construct arbitrary queries.

Thus, we categorize the attack model into three levels as follows:

- Level 1: Attackers can only observe the encrypted data stored in the cloud. This follows the well-known ciphertext-only attack model [20].
- Level 2: In addition to the encrypted data stored in the cloud, attackers are able to get a set of biometric traits in the database D but do not know the corresponding ciphertexts in the database C , which is similar to the known-candidate attack model [21].
- Level 3: Besides all the abilities in level-2, attackers in level-3 can be valid users. Thus, attackers can forge as many identification queries as possible and obtain the corresponding ciphertexts. This attack follows the known-plaintext attack model [20].

A biometric identification scheme is secure if it can resist the level-α($\alpha \in \{1, 2, 3\}$) attack. Note that that if the proposed scheme can resist level-2 and level-3 attacks, it does not mean that the attacker can both be the valid user and observe some plaintexts of the biometric database simultaneously. This sophisticated attack is too strong and no effective methods is designed to defend against this kind of attack [14]. In this paper, we focus on the collusion attack between a malicious user and the cloud server. The relationship between the plaintexts of the biometric database and the ciphertexts is not known to the attacker, which is similar to the attack model proposed in [14].

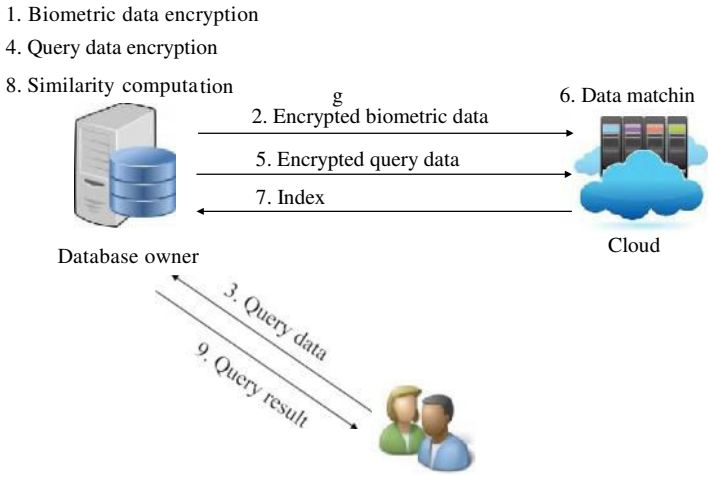


FIGURE 1. System model.

C. DESIGN GOALS

In order to achieve practicality, both security and efficiency are considered in the proposed scheme. To be more specific, design goals of the proposed scheme are described as follows:

- Efficiency: Computational costs should be as low as possible at both the database owner side and the user side. To gain high efficiency, most biometric identification operations should be executed in the cloud.
- Security: During the identification process, the privacy of biometric data should be protected. Attackers and the semi-honest cloud should learn nothing about the sensitive information.

D. NOTATIONS

Here, we list the main notations used in the remaining section as follows.

- b_i – the i-th sample FingerCode, denoted as an n-dimensional vector $b_i = [b_{i1}, b_{i2}, \cdots b_{in}]$.
- B_i – the extended sample FingerCode of b_i, denoted as an $(n+1)$-dimensional vector $B_i = [b_{i1}, b_{i2}, \cdots b_{i(n+1)}]$, where $b_{i(n+1)} = -0.5(b_{i1}^2 + b_{i2}^2 + \cdots + b_{in}^2)$.
- b_c – the query FingerCode, denoted as an n-dimensional vector $b_c = [b_{c1}, b_{c2}, \cdots b_{cn}]$.
- B_c – the extended query FingerCode of b_c, denoted as an $(n+1)$-dimensional vector $B_c = [b_{c1}, b_{c2}, \cdots b_{c(n+1)}]$, where $b_{c(n+1)} = 1$.
- W – the secret keys collection, denoted as $W = (M_1, M_2, M_3, H, R)$, where M_1, M_2 and M_3 are $(n+1) \times (n+1)$ invertible matrices, and H, R are $(n+1)$-dimensional row vectors.
- I_i – the searchable index associated with the i-th sample FingerCode b_i.
- Γ – the query FingerCodes collection constructed by the attacker, denoted as $\Gamma = (b_1, b_2, \cdots b_{t+1})$.
- B_i – the i-th extended query FingerCode constructed by the attacker, denoted as $B_i = [b_{i1}, b_{i2}, \cdots b_{i(n+1)}]$, where $b_{i(n+1)} = 1$.

III. SECURITY ANALYSIS OF YUAN AND YU'S SCHEME

In this section, we firstly describe Yuan and Yu's scheme and then give the security analysis about their scheme. To facilitate understanding of the scheme, we use $*$ to denote the elements multiplication operations, and use $\times$ to denote the matrices or vectors multiplication operations.

A. YUAN AND YU'S SCHEME

Step 1: The database owner randomly generates an $(n+1) \times (n+1)$ matrix A where $H \times A^T = 1$ and A_i is a row vector in A, $1 \le i \le (n+1)$.

Then, the database owner generates a corresponding matrix $D_i = [A^1 * b_{i1}, A^2 * b_{i2}, \cdots A^{n+1} * b_{i(n+1)}]$ to hide B_i.

After that, the database owner performs the following operations:

$$C_i = M_1 \times D_i \times M_2, \tag{1}$$

$$C_h = H \times M_1^{-1}, \tag{2}$$

$$C_r = M_3^{-1} \times R^T. \tag{3}$$

Subsequently, the database owner uploads (C_i, C_h, C_r, I_i) to the cloud, where I_i is the index of B_i.

Step 2: After Step 1 is executed, the cloud has stored many tuples in its database C. When a user requests to identify his/her identity, he/she extends b_i and then submits

the extended query B_i to the database owner. On receiving the request from the user, the database owner generates a random $(n+1)\times(n+1)$ matrix E such that $E_i \times R^T = 1$, where E_i is a row vector in matrix E and $1 \le i \le (n+1)$. The database owner then generates a corresponding matrix $F_c = [E^{-1}_1 * b_{c1}, E^{-1}_2 * b_{c2}, \cdots E^{-1}_{n+1} * b_{c(n+1)}]^T$ to hide the query FingerCode B_c. The Database owner then performs the following operations:

$$C_f = M_2^{-1} \times F_c \times M_3. \tag{4}$$

Then, the database owner uploads C_f to the cloud.

Step 3: On receiving C_f, the cloud begins to search for the best match. Specifically, the cloud computes $P_i = C_h \times C_i \times C_f \times C_r$ for all encrypted biometric database to compare the Euclidean distances between b_c and b_i. Other details are eliminated since they are irrelevant for the security analysis we will describe.

B. SECURITY ANALYSIS OF YUAN AND YU'S SCHEME

In level-3 attack, an attacker has the ability to select query FingerCodes Γ of his/her interest as inputs and then tries to recover the privacy of B_i. Specifically, the attacker can compute the secret key M_2 by performing the following equation:

$$\begin{aligned} C_f \times C_r &= M_2^{-1} \times F_c \times M_3 \times M_3^{-1} \times R^T \\ &= M_2^{-1} \times F_c \times R^T \\ &= M_2^{-1} \times B_c^T. \end{aligned} \tag{5}$$

In equation 5, C_f is an $(n+1)\times(n+1)$ matrix and C_r is an $(n+1)$-dimensional vector which are both known to the attacker. B_c is an $(n+1)$-dimensional vector which can be constructed by the attacker. M_2^{-1} is one of the secret keys which is an $(n+1)\times(n+1)$ matrix but unknown to the attacker. Let S be $C \times C_f$. To recover M_2^{-1}, t query FingerCodes $\Gamma = [b_1, b_2, \cdots\cdots b_t]$ which are extended to $[\tilde{B}^T_1, \tilde{B}^T_2, \cdots\cdots \tilde{B}^T_t]$ can be constructed, such that

$$[S_1, S_2, \cdots S_t] = M_2^{-1} \times [\tilde{B}^T_1, \tilde{B}^T_2 \cdots \tilde{B}^T_t]. \tag{6}$$

There are $(n+1)\times t$ known elements in $[S_1, S_2, \cdots\cdots S_t]$ and $(n+1)\times t$ known elements in $[\tilde{B}^T_1, \tilde{B}^T_2, \cdots\cdots \tilde{B}^T_t]$, M_2^{-1} is a matrix with $(n+1)\times(n+1)$ unknown elements. Suppose $M_2 = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1(n+1)} \\ \vdots & \vdots & \ddots & \vdots \\ q_{(n+1)1} & q_{(n+1)2} & \cdots & q_{(n+1)(n+1)} \end{bmatrix}$, we will show how to recover M_2^{-1} by constructing special FingerCodes.

For the first row vector $q_1 = [q_{11}, q_{12}, \cdots, q_{1(n+1)}]$ in M_2^{-1}, the adversary constructs two special vectors as $\tilde{B}_1^T = [1, 0, \cdots, -0.5]$, and $\tilde{B}_2^T = [2, 0, \cdots, -2]$. Then, the attacker can compute as follows:

$$\begin{aligned} 1 * q_{11} - 0.5 * q_{1(n+1)} &= S_{11}, \\ 2 * q_{11} - 2 * q_{1(n+1)} &= S_{21}. \end{aligned} \tag{7}$$

From equation 7, it is easy to compute q_{11} and $q_{1(n+1)}$. Following the same analysis, the attacker can obtain all the elements in M_2^{-1} by constructing other special vectors.

After recovering M_2^{-1}, the attacker can compute the biometric data as follows:

$$\begin{aligned} C_h \times C_i &= H \times M^{-1} \times M \times D_i \times M_2 \\ &= H \times D_i \times M_2 \\ &= B_i \times M_2. \end{aligned} \tag{8}$$

In equation 8, C_h and C_i are known by the attacker. M_2 is the secret key which is recovered by the above foregoing. Therefore, the attacker can recover B_i.

IV. A NOVEL BIOMETRIC IDENTIFICATION SCHEME

In this section, we show the details of the proposed biometric identification scheme.

A. OVERVIEW

We construct a novel biometric identification scheme to address the weakness of Yuan and Yu's scheme [13]. To achieve a higher level of privacy protection, a new retrieval way is constructed to resist the level-3 attack. Moreover, we also reconstruct the ciphertext to reduce the amount of uploaded data and improve the efficiency both in the preparation and identification procedures.

In the remaining part of this section, we will introduce the preparation process and the identification process.

B. PREPARATION PROCESS

In the preparation process, b_i is the i-th sample feature vector derived from the fingerprint image using a feature extraction algorithm [19]. To be more specific, b_i is an n-dimensional vector with l bits of each element where $n = 640$ and $l = 8$.

For ease of identification, b_i is extended by adding an $(n+1)$-th element as B_i. Then, the database owner encrypts B_i with the secret key M_1 as follows:

$$C_i = B_i \times M_1. \tag{9}$$

The database owner further performs the following operation:

$$C = M^{-1} \times H. \tag{10}$$

Each FingerCode B_i is associated with an index I_i. After execute the encryption operations, the database owner uploads (C_i, C_h, I_i) to the cloud.

C. IDENTIFICATION PROCESS

The identification process includes the following steps:

Step 1: When a user has a query fingerprint to be identified, he/she first gets the query FingerCode b_c derived from the query fingerprint image. The FingerCode b_c is also an n-dimensional vector. Then, the user sends b_c to the database owner.

Step 2: After receiving b_c, the database owner extends b_c to B_c by adding an $(n+1)$-th element equals to 1. Then the database owner randomly generates an $(n+1) \times (n+1)$ matrix E. The i-th row vector $E_i = [E_{i1}, E_{i2}, \cdots E_{i(n+1)}]$ is set as a random vector, where the $(n+1)$-th element is $(1 - \sum_{j=1}^{n} E_{ij} * H_j)/H_{n+1}$, $1 \leq i \leq (n+1)$. After that, the database owner performs the following computation to hide B_c:

$$F_c = [E^T_1 * b_{c1}, E^T_2 * b_{c2}, \cdots E^T_{(n+1)} * b_{c(n+1)}]^T. \quad (11)$$

To securely send F_c to the cloud, the database owner needs to encrypt F_c with the secret keys and a random integer $r (r > 0)$. The computation is performed as follows:

$$C_f = M^{-1}_1 \times r \times F_c \times M_2. \quad (12)$$

Then, the database owner sends C_f to the cloud for identification.

Step 3: After receiving C_f from the database owner, the cloud begins to search the FingerCode which has the minimum Euclidean distance with the query FingerCode B_c. P_i denotes the relative distance between B_i and B_c as follows:

$$\begin{aligned} P_i &= C_i \times C_f \times C_h \\ &= B_i \times M_1 \times M^{-1}_1 \times r \times F_c \times M_2 \times M^{-1}_2 \times H^T \\ &= B_i \times r \times F_c \times H^T \\ &= \sum_{j=1}^{n+1} r * b_{ij} * b_{cj}. \end{aligned} \quad (13)$$

In equation 13, the computation result is an integer, which can be used to compare two FingerCodes. For example, to compare the query b_c with two FingerCodes, say b_i and b_z, the cloud computes P_i and P_z, and performs the following operation, where $1 \leq i, z \leq t$, $i \neq z$:

$$\begin{aligned} P_i - P_z &= \sum_{j=1}^{n+1} r * b_{ij} * b_{cj} - \sum_{j=1}^{n+1} r * b_{zj} * b_{cj} \\ &= \left(\sum_{j=1}^{n} r * b_{ij} * b_{cj} - 0.5 \sum_{j=1}^{n} r * b^2_{ij} \right) \\ &\quad - \left(\sum_{j=1}^{n} r * b_{zj} * b_{cj} - 0.5 \sum_{j=1}^{n} r * b^2_{zj} \right) \\ &= 0.5r(dist^2_x - dist^2_{ic}). \end{aligned} \quad (14)$$

As shown in equation 14, if $P_i - P_z > 0$, the cloud learns that b_i matches the query FingerCode much better than b_z. After repeating the operations for the encrypted FingerCode database C in the cloud, the ciphertext C_i which has the minimum Euclidean distance with b_c can be found. The cloud further gets the corresponding index I_i according to the tuple (C_i, C_h, I_i) and sends it back to the database owner.

Step 4: After receiving the index I_i, the database owner gets the corresponding sample FingerCode b_i in the database D and calculates the accurate Euclidean distance between b_i and b_c as $dist_{ic} = \sum_{j=1}^{n} (b_{ij} - b_{cj})^2$. Then, the database owner compares the Euclidean distance with the standard threshold. If the distance is less than the threshold value, the query is identified. Otherwise, the identification fails.

Step 5: Finally, the database owner returns the identification result to the user.

V. SECURITY ANALYSIS

In this part, we first prove that our scheme is secure under level-2 and level-3 attacks, and then we will show the proposed scheme can resist the attack proposed by Zhu et al [18].

A. SECURITY ANALYSIS UNDER LEVEL-2 ATTACK

According to the attack scenario 2, an attacker can obtain some plaintexts of the biometric database, but does not know the corresponding ciphertexts.

We consider C_i which is obtained by multiplying B_i and M_1. Since the mapping relationship between B_i and C_i is not known, it is impossible for the attacker to compute B_i and M_1.

B. SECURITY ANALYSIS UNDER LEVEL-3 ATTACK

In the level-3 attack, besides the knowledge of encrypted data in the cloud, the attacker can forge a large number of query FingerCodes Γ as inputs. In the following, we will show the proposed scheme is secure by proving that the secret keys cannot be recovered.

When colluding with the cloud, the attacker gets C_f and C_h, and then performs the following operation:

$$\begin{aligned} C_f \times C_h &= M^{-1}_1 \times r \times F_c \times M_2 \times M^{-1}_2 \times H^T \\ &= M^{-1}_1 \times r \times F_c \times H^T \\ &= M^{-1}_1 \times r \times B^T_i. \end{aligned} \quad (15)$$

In equation 15, since r is a positive random integer in identification process, the attacker cannot compute the secret key M^{-1}_1 directly.

Pretending a valid user, the attacker can construct t query FingerCodes $\Gamma = [\tilde{b}_1, \tilde{b}_2, \cdots \tilde{b}_t]$ extended as $[\tilde{B}_1, \tilde{B}_2, \cdots \tilde{B}_t]$ for identification, which introduces a set of positive random values r_j and C_{fj}, $1 \leq j \leq t$. Let $\tilde{P}_j$ be the value of $C_{fj} \times C_h$. The attacker computes $\tilde{P}_j$ as follows:

$$\tilde{P}_j = M^{-1}_1 \times r_j \times \tilde{B}^T_j. \quad (16)$$

After constructing t equations, we have:

$$\begin{aligned} \tilde{P} &= M^{-1}_1 \times [\tilde{B}^T_1, \tilde{B}^T_2, \cdots \tilde{B}^T_t] \times \begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_t \end{bmatrix} \\ &= M^{-1}_1 \times \tilde{B} \times R. \end{aligned} \quad (17)$$

Here $[\tilde{B}_1^T, \tilde{B}_2^T, \cdots \tilde{B}_t^T]$ is denoted as B, $\begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_t \end{bmatrix}$ is denoted as R. In this equation, P is an $(n+1) \times t$ matrix known to the attacker, $\tilde{B}$ is an $(n+1) \times t$ matrix constructed by the attacker, R is an $t \times t$ matrix, since r_j is a random positive integer, it is unknown to the attacker.

We then demonstrate that the attacker cannot recover M_1 according to **Theorem 1.**

Theorem 1. Assume after t equationsare constructed, M_1 cannot be computed in $\tilde{P} = M_1^{-1} \times \tilde{B} \times R$. When $(t+1)$ equationsare constructed, the following equation holds, and M_1 cannot be recovered.

$$\tilde{P}|\tilde{P}_{t+1} = M_1^{-1} \times [\tilde{B}|\tilde{B}^T_{t+1}] \times \begin{bmatrix} R & 0 \\ 0 & r_{t+1} \end{bmatrix}. \quad (18)$$

Proof. This theorem is proven with *the inductive method.* When $t = 1$, M_1 cannot be computed in equation 16. Assume the equation 17 holds, where $(t > 1)$. When $(t+1)$ query FingerCodes are constructed, we obtain:

$$\tilde{P}, \tilde{P}_{t+1} = [M_1^{-1} \times \tilde{B}, M_1^{-1} \times \tilde{B}^T_{t+1}] \times \begin{bmatrix} R & 0 \\ 0 & r_{t+1} \end{bmatrix}. \quad (19)$$

For $(t+1)$-th query FingerCode $\tilde{B}_{t+1}$, we have

$$\tilde{P}_{t+1} = M_1^{-1} \times \tilde{B}^T_{t+1} \times r_{t+1}. \quad (20)$$

From equation 20, we have

$$\tilde{B}_{t+1} \times (M_1^{-1})^T = (r_t^{-1})^T \times \tilde{P}^T_{t+1}. \quad (21)$$

Let $(d_1, d_2, \cdots d_{n+1})$ be the vector $(r^{-1})^T \times \tilde{P}^T_{t+1}$ where $d_j = (r_t^{-1})^T \times P_{(t+1)j}$. Let $(M_1^{-1})^T = (m_1^T, m_2^T, \cdots, m_{n+1}^T)$, where m_j denotes a row vector in M_1^{-1}, $1 \le j \le (n+1)$. The following equations hold:

$$\tilde{B}_{t+1} \times (m_1^T, m_2^T, \cdots m_{n+1}^T)^T = (d_1, d_2, \cdots, d_{n+1}), \quad (22)$$

$$\tilde{B}^T_j \times m_j = d_j \quad (23)$$

Equation 23 is a typical non-linear nonhomogeneous equation. Since the rank of $\tilde{B}_{t+1}$ is $r(\tilde{B}_{t+1})$, we assume the result is $\alpha_1\beta_1 + \alpha_2\beta_2 + \cdots + \alpha_{(n+1-r(\tilde{B}_{t+1}))}\beta_{n+1-r(\tilde{B}_{t+1})}$. We further state the special solution of equation 23 is β^* which satisfies the formula $\tilde{B}_{t+1} \times m_j = d_j$. Because $d_j = (r_{t+1})_j \times \tilde{P}^T_{t+1} \times (r^{-1})^T$, β^* is included in the special solution of m_j^T. For m_j^T in matrix $(M_1^{-1})^T$, the particular solution of m_j^T is $\alpha_1\beta_1 + \alpha_2\beta_2 + \cdots + \alpha_{(n+1-r(\tilde{B}_{t+1}))}\beta_{n+1-r(\tilde{B}_{t+1})} + \beta^*$. Since r is a random integer, the special solution β^* is uncertain as well, which means the attacker cannot derive the exact particular solution for m_j^T in $(M_1^{-1})^T$. □

Therefore, when $(t+1)$ query FingerCodes are constructed, the secret key M_1 cannot be computed by the attacker as well.

As discussed above, the attacker cannot recover the secret key even if he is a malicious user. Therefore, the attacker cannot recover the biometric data as well.

Moreover, we compare our scheme with the schemes proposed in [13] and [14]. According to Table 1, other schemes have some weaknesses, while our scheme is secure under all the three level attacks

C. SECURITY ANALYSIS UNDER THE ATTACK PROPOSED BY ZHU ET AL.

Zhu et al. [18] showed an attack for Yuan and Yu's scheme. In their attack, the attacker observes the cloud and gets the values of relative distance. According to the equation 1, 2, 3, 4, the relative distance in Yuan and Yu's scheme can be computed as follows:

$$\begin{aligned} P_i &= C_h \times C_i \times C_f \times C_r \\ &= H \times M_1^{-1} \times M_1 \times D_i \times M_2 \\ &\quad \times M_3^{-1} \times F_c \times M_3 \times M_2^{-1} \times R^T \\ &= H \times D_i \times F_c \times R^T \\ &= \sum_{j=1}^{n+1} b_{ij} * b_{cj} \\ &= B_i \times B_c^T. \end{aligned} \quad (24)$$

As shown in equation 24, P_i is an integer which the attacker can get in the cloud, B_c is the extended query FingerCode which can be constructed by the attacker pretending to be a user. B_i is the extended sample FingerCode which is sensitive and should not be leaked. To recover B_i, the attacker can construct t query FingerCodes $\Gamma = [\tilde{b}_1, \tilde{b}_2, \ldots \tilde{b}_t]$ extended as $[B_1, B_2, \cdots B_t]$ for identification. P_{ij} denotes the relative distance between the sample FingerCode B_i and the query FingerCode $\tilde{B}_j$ where $1 \le j \le t$. Then, the attacker has:

$$[\tilde{P}_{i1}, \tilde{P}_{i2}, \cdots \tilde{P}_{it}] = [b_{i1}, b_{i2}, \cdots b_{i(n+1)}] \times [\tilde{B}^T_1 \tilde{B}^T_2 \cdots \tilde{B}^T_t]. \quad (25)$$

In this equation, $\tilde{P}_{ij}$ and $\tilde{B}_j$ are known to the attacker. For each element in B_i, it can be recovered if t equations are built, where $t > (n+1)$.

Then, we demonstrate the proposed scheme is secure under the attack proposed by Zhu et al. In the proposed scheme, $\tilde{P}_{ij}$ is set as the relative distance between B_i and $\tilde{B}_j$.

$$\begin{aligned} \tilde{P}_{ij} &= C_i \times C_{fj} \times C_{h_j} \\ &= r_j \times B_i \times \tilde{B}^T. \end{aligned} \quad (26)$$

r_j is the j-th positive random integer in t identification processes. The attacker constructs t query FingerCodes and gets the equation as follows:

TABLE 1. Security comparison with other schemes.

Schemes	Level 1 attack	Level 2 attack	Level 3 attack
Yuan and Yu's scheme [13]	Yes	Yes	No
Wang et al.'s scheme [14]	Yes	Yes	No
Our scheme	Yes	Yes	Yes

$$[\tilde{P}_{i1}, \tilde{P}_{i2}, \cdots \tilde{P}_{it}] = [b_{i1}, b_{i2}, \cdots b_{i(n+1)}] \times [r_1\tilde{B}_l^T, r_2\tilde{B}_2^T, \cdots r_t\tilde{B}_t^T]$$

$$= [b_{i1}, b_{i2}, \cdots b_{i(n+1)}] \times [B_1, B_2, \cdots B_t] \times \begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_t \end{bmatrix} \quad (27)$$

In this equation, r_j is a positive random integer which is unknown to the attacker. For every element in B_i, after t computations, the attacker can only get the value of $r_j * b_{iq}$ where $t > (n+1)$, $1 \le q \le (n+1)$. For the reason that r_j is a random integer, $r_j * b_{iq}$ is also unexpected which means the attacker cannot acquire B_i. Thus, the proposed scheme can resist the attack proposed by Zhu et al.

VI. PERFORMANCE ANALYSIS

To evaluate the performance of the proposed scheme, we implement a cloud-based privacy-preserving fingerprint identification system. For the cloud, we use 2 nodes with 6-core 2.10 GHz Intel Xeous CPU and 32GB memory. We utilize a laptop with an Intel Core 2.40 GHz CPU and 8G. Similar to [13] and [14], the query FingerCodes are randomly selected from the database which is constructed with random 640-entry vectors.

A. COMPLEXITY ANALYSIS

Table 2 summarizes the computation and communication costs on the data owner side, cloud server and users in our scheme and the schemes in [13] and [14]. In this work, each matrix multiplication costs $O(n^3)$, where n denotes the dimension of a FingerCode, and the sorting cost of fuzzy Euclidean distances has time complexity of $O(m \log m)$. As illustrated in Table 2, our scheme has lower complexities in the preparation phase. That is, more computation and bandwidth costs can be saved for the database owner. In the identification phase, the computation complexity of our scheme is lower than that in [14]. The reason is that our scheme performs vector-matrix multiplication operations to find the close match, while [14] needs to execute matrix-matrix multiplication operations. Although the complexity of our scheme is the same as that in [13], we emphasize that [13] sacrifices the substantial security to achieve such fast computation of P_i. Moreover, our scheme executes fewer multiplication operations, and thus obtains better performance.

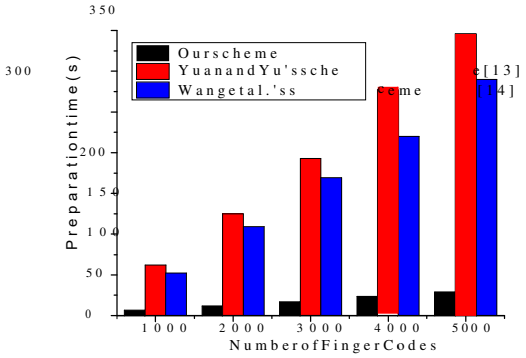


FIGURE 2. Time costs in the preparation phase.

B. EXPERIMENTAL EVALUATION

Preparation phase. Fig. 2 and Fig. 3 show the computation and communication costs in the preparation phase with the number of FingerCodes varying from 1000 to 5000. As shown in Fig.2, in our scheme, registering 5000 FingerCodes needs 29.37s, which can save about 88.85% and 90.58% time cost compared with [13] and [14] respectively. The reason is when encrypting a sample FingerCode, in our scheme, only one matrix is needed which leads to fewer matrix multiplication operations. Fig. 3 shows the bandwidth costs of the three schemes. Since the data outsourced to the cloud is in the form of vectors in comparison with matrices in the other two schemes, the communication cost in our scheme is much less than [13], [14].

Identification phase. Fig.4 and Fig. 5 show the computation and communication costs in the identification phase with the number of FingerCodes ranges from 1000 to 5000. As demonstrated in Fig. 4, all schemes grow linearly as the size of database increases. As in our scheme fewer matrix multiplication operations are used than [13], it can save about 56% time cost. Compared with [14], the identification time can be saved as much as 84.75%, since the vector-matrix multiplication rather than the matrix-matrix multiplication operation is executed. The bandwidth costs of the three schemes, as shown in Fig. 5, are almost the same. The reason is that all schemes need to transmit a matrix in the identification phase.

TABLE 2. A summary of complexity costs. In the table, m denotes the number of FingerCodes in the biometric database; n m.

		Phases	Yuan and Yu's scheme [13]	Wang et al.'s scheme [14]	Our scheme
Computation	Database owner	Preparation	$O(mn^3)$	$O(mn^3)$	$O(mn^2)$
		Identification	$O(n^3)$	$O(n^3)$	$O(n^3)$
		Retrieval	$O(n)$	$O(n)$	$O(n)$
	Cloud server	Identification	$O(mn^2 + m \log m)$	$O(mn^3 + m \log m)$	$O(mn^2 + m \log m)$
	User	Identification	/	/	/
Communication	Database owner	Preparation	$O(mn^2)$	$O(mn^2)$	$O(mn)$
		Identification	$O(n^2)$	$O(n^2)$	$O(n^2)$
		Retrieval	$O(1))$	$O(1)$	$O(1)$
	Cloud server	Identification	/	/	/
		Retrieval	$O(1)$	$O(1)$	$O(1)$
	User	Identification	$O(1)$	$O(1)$	$O(1)$

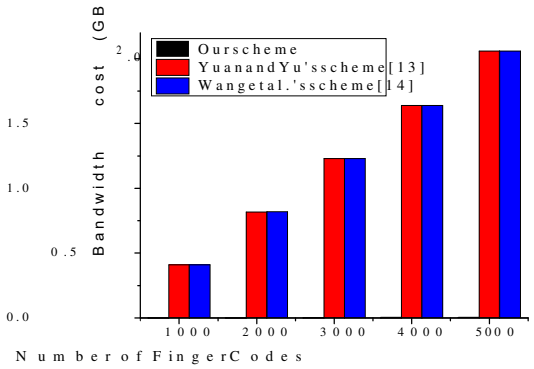


FIGURE 3. Bandwidth costs in the preparation phase.

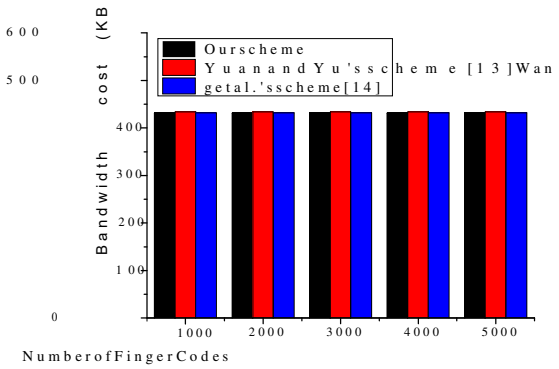


FIGURE 5. Bandwidth costs in the identification phase.

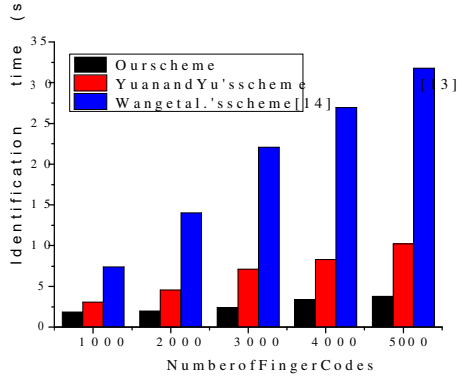


FIGURE 4. Time costs in the identification phase.

VII. RELATED WORKS

Related works on privacy-preserving biometric identification are provided in this section. Recently, some efficient biometric identification schemes have been proposed. Wang and Hatzinakos proposed a privacy-preserving face recognition scheme [22]. Specifically, a face recognition method is designed by measuring the similarity between sorted index numbers vectors. Wong and Kim [23] proposed a privacy-preserving biometric matching protocol for iris codes verification. In their protocol, it is computationally infeasible for a malicious user to impersonate as an honest user. Barni et al. [10] presented a FingerCode identification protocol based on the Homomorphic Encryption technique. However, all distances are computed between the query and sample Fingercodes in the database, which introduces too much burden as the size of fingerprints increases. To improve the efficiency, Evans et al. [12] proposed a novel protocol which reduces the identification time. They used an improved Homomorphic encryption algorithm to compute the Euclidean distance and designed novel garbled circuits to find the minimum distance. By exploiting a backtracking protocol, the best match FingerCode can be found. However, in [12], the whole encrypted database has to be transmitted to the user from the database server. Wong et al. [24] proposed an identification scheme

based on kNN to achieve secure search in the encrypted database. However, their scheme assumes that there is no collusion between the client side and cloud server side. Yuan and Yu [13] proposed an efficient privacy-preserving biomet- ric identification scheme. However, Zhu et al. [18] pointed out their protocol can be broken if a malicious user colludes with the cloud server in the identification process. Based on [13], Wang et al. presented a privacy-preserving biometric identification scheme in [14] which introduced random di- agonal matrices, named CloudBI-II. However, their scheme has been proven insecure in [15], [16]. Recently, Zhang et al. [17] proposed an efficient privacy-preserving biometric identification scheme by using perturbed terms.

VIII. CONCLUSION

In this paper, we proposed a novel privacy-preserving biometric identification scheme in the cloud computing. To realize the efficiency and secure requirements, we have designed a new encryption algorithm and cloud authentication certification. The detailed analysis shows it can resist the potential attacks. Besides, through performance evaluations, we further demonstrated the proposed scheme meets the efficiency need well.

REFERENCES

[1] A. Jain, L. Hong and S. Pankanti, “Biometric identification,” Communications of the ACM, vol. 43, no. 2, pp. 90-98, 2000.
[2] R. Allen, P. Sankar and S. Prabhakar, “Fingerprint identification technology,” Biometric Systems, pp. 22-61, 2005.
[3] J. de Mira, H. Neto, E. Neves, et al., “Biometric-oriented Iris Identification Based on Mathematical Morphology,” Journal of Signal Processing Systems, vol. 80, no. 2, pp. 181-195, 2015.
[4] S. Romdhani, V. Blanz and T. Vetter, “Face identification by fitting a 3d morphable model using linear shape and texture error functions,” in European Conference on Computer Vision, pp. 3-19, 2002.
[5] Y. Xiao, V. Rayi, B. Sun, X. Du, F. Hu, and M. Galloway, “A survey of key management schemes in wireless sensor networks,” Journal of Computer Communications, vol. 30, no. 11-12, pp. 2314-2341, 2007.
[6] X. Du, Y. Xiao, M. Guizani, and H. H. Chen, “An effective key management scheme for heterogeneous sensor networks,” Ad Hoc Networks, vol. 5, no. 1, pp. 24-34, 2007.
[7] X. Du and H. H. Chen, “Security in wireless sensor networks,” IEEE Wireless Communications Magazine, vol. 15, no. 4, pp. 60-66, 2008.
[8] X. Hei, and X. Du, “Biometric-based two-level secure access control for implantable medical devices during emergency,” in Proc. of IEEE INFOCOM 2011, pp. 346-350, 2011.
[9] X. Hei, X. Du, J. Wu, and F. Hu, “Defending resource depletion attacks on implantable medical devices,” in Proc. of IEEE GLOBECOM 2010, pp. 1-5, 2010.
[10] M. Barni, T. Bianchi, D. Catalano, et al., “Privacy-preserving fingercode authentication,” in Proceedings of the 12th ACM workshop on Multimedia and security, pp. 231-240, 2010.
[11] M. Osadchy, B. Pinkas, A. Jarrous, et al., “SCiFI-a system for secure face identification,” in Security and Privacy (SP), 2010 IEEE Symposium on, pp. 239-254, 2010.
[12] D. Evans, Y. Huang, J. Katz, et al., “Efficient privacy-preserving biometric identification,” in Proceedings of the 17th conference Network and Distributed System Security Symposium, NDSS, 2011.
[13] J. Yuan and S. Yu, “Efficient privacy-preserving biometric identification in cloud computing,” in Proc. of IEEE INFOCOM 2013, pp. 2652-2660,2013.
[14] Q. Wang, S. Hu, K. Ren, et al., “CloudBI: Practical privacy-preserving outsourcing of biometric identification in the cloud,” in European Symposium on Research in Computer Security, pp. 186-205, 2015.
[15] Y. Zhu, Z. Wang and J. Wang, “Collusion-resisting secure nearest neighbor query over encrypted data in cloud,” In Quality of Service (IWQoS), 2016 IEEE/ACM 24th International Symposium on, pp. 1-6, 2016.

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

Cyberbullying Detection based on Semantic-Enhanced Marginalized Denoising Auto-Encoder

L.Shruthi
M.Tech. SCHOLAR,
CSE Department
Malla Reddy College of Engineering

Abstract—As a side effect of increasingly popular social media, cyberbullying has emerged as a serious problem afflicting children, adolescents and young adults. Machine learning techniques make automatic detection of bullying messages in social media possi ble, and this could help to construct a healthy and safe social media environment. In this meaningful research area, one critical issue is robust and discriminative numerical representation learning of text messages. In this paper, we propose a new representation learning method to tackle this problem. Our method named Semantic-Enhanced Marginalized Denoising Auto-Encoder (smSDA) is developed via semantic extension of the popular deep learning model stacked denoising autoencoder. The semantic extension consists of semantic dropout noise and sparsity constraints, where the semantic dropout noise is designed based on domain knowledge and the word embedding technique. Our proposed method is able to exploit the hidden feature structure of bullying information and learn a robust and discriminative representation of text. Comprehensive experiments on two public cyberbullying corpora (*Twitter* and *MySpace*) are conducted, and the results show that our proposed approaches outperform other baseline text representation learning methods.

Index Terms—Cyberbullying Detection, Text Mining, Representation Learning, Stacked Denoising Autoencoders, Word Embedding

1 INTRODUCTION

SOCIAL Media, as defined in [1], is ‘’a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content.‘’ Via social media, people can enjoy enormous information, convenient communication experience and so on. However, social me- dia may have some side effects such as cyberbullying, which may have negative impacts on the life of people, especially
children and teenagers.

Cyberbullying can be defined as aggressive, intentional actions performed by an individual or a group of people via digital communication methods such as sending messages and posting comments against a victim. Different from tra- ditional bullying that usually occurs at school during face- to-face communication, cyberbullying on social media can take place anywhere at any time. For bullies, they are free to hurt their peers’ feelings because they do not need to face someone and can hide behind the Internet. For victims, they are easily exposed to harassment since all of us, especially youth, are constantly connected to Internet or social media. As reported in [2], cyberbullying victimization rate ranges from 10% to 40%. In the United States, approximately 43% of teenagers were ever bullied on social media [3]. The same as traditional bullying, cyberbullying has negative, insidious and sweeping impacts on children [4], [5], [6]. The outcomes for victims under cyberbullying may even be tragic such as the occurrence of self-injurious behaviour or suicides.

•

One way to address the cyberbullying problem is to automatically detect and promptlyreport bullying messages so that proper measures can be taken to prevent possi- ble tragedies. Previous works on computational studies of bullying have shown that natural language processing and machine learning are powerful tools to study bullying [7], [8]. Cyberbullying detection can be formulated as a supervised learning problem. A classifier is first trained on a cyberbullying corpus labeled by humans, and the learned classifier is then used to recognize a bullying message. Three kinds of information including text, user demography, and social network features are often used in cyberbullying detection [9]. Since the text content is the most reliable, our work here focuses on text-based cyberbullying detection. In the text-based cyberbullying detection, the first and also critical step is the numerical representation learning for text messages. In fact, representation learning of text is extensively studied in text mining, information retrieval and natural language processing (NLP). Bag-of-words (BoW) model is one commonly used model that each dimension corresponds to a term. Latent Semantic Analysis (LSA) and topic models are another popular text representation models, which are both based on BoW models. By mapping text units into fixed-length vectors, the learned represen- tation can be further processed for numerous language processing tasks. Therefore, the useful representation should discover the meaning behind text units. In cyberbullying detection, the numerical representation for Internet mes- sages should be robust and discriminative. Since messages on social media are often very short and contain a lot of informal language and misspellings, robust representations for these messages are required to reduce their ambigui-
ty. Even worse, the lack of sufficient high-quality training

ISBN 978-93-88808-61-3

data, i.e., data sparsity make the issue more challenging. Firstly, labeling data is labor intensive and time consuming. Secondly, cyberbullying is hard to describe and judge from a third view due to its intrinsic ambiguities. Thirdly, due to protection of Internet users and privacy issues, only a small portion of messages are left on the Internet, and most bullying posts are deleted. As a result, the trained classifier may not generalize well on testing messages that contain nonactivated but discriminative features. The goal of this present study is to develop methods that can learn robust and discriminative representations to tackle the above prob- lems in cyberbullying detection.

Some approaches have been proposed to tackle these problems by incorporating expert knowledge into feature learning. Yin et.al proposed to combine BoW features, senti- ment features and contextual features to train a support vec- tor machine for online harassment detection [10]. Dinakar et.al utilized label specific features to extend the general features, where the label specific features are learned by Linear Discriminative Analysis [11]. In addition, common sense knowledge was also applied. Nahar et.al presented a weighted TF-IDF scheme via scaling bullying-like features by a factor of two [12]. Besides content-based information, Maral et.al proposed to apply users' information, such as gender and history messages, and context information as extra features [13], [14]. But a major limitation of these approaches is that the learned feature space still relies on the BoW assumption and may not be robust. In addition, the performance of these approaches rely on the quality of hand-crafted features, which require extensive domain knowledge.

In this paper, we investigate one deep learning method named stacked denoising autoencoder (SDA) [15]. SDA stacks several denoising autoencoders and concatenates the output of each layer as the learned representation. Each denoising autoencoder in SDA is trained to recover the input data from a corrupted version of it. The input is corrupted by randomly setting some of the input to zero, which is called dropout noise. This denoising process helps the autoencoders to learn robust representation. In addition, each autoencoder layer is intended to learn an increasingly abstract representation of the input [16]. In this paper, we develop a new text representation model based on a variant of SDA: marginalized stacked denoising autoencoders (mS- DA) [17], which adopts linear instead of nonlinear projection to accelerate training and marginalizes infinite noise distri- bution in order to learn more robust representations. We utilize semantic information to expand mSDA and develop Semantic-enhanced Marginalized Stacked Denoising Au- toencoders (smSDA). The semantic information consists of bullying words. An automatic extraction of bullying words based on word embeddings is proposed so that the involved human labor can be reduced. During training of smSDA, we attempt to reconstruct bullying features from other normal words by discovering the latent structure, i.e. correlation, between bullying and normal words. The intuition behind this idea is that some bullying messages do not contain bullying words. The correlation information discovered by smSDA helps to reconstruct bullying features from normal words, and this in turn facilitates detection of bullying messages without containing bullying words. For example, there is a strong correlation between bullying word *fuck* and normal word *off* since they often occur together. If bullying messages do not contain such obvious bullying features, such as *fuck* is often misspelled as *fck*, the correlation may help to reconstruct the bullying features from normal ones so that the bullying message can be detected. It should be noted that introducing dropout noise has the effects of enlarging the size of the dataset, including training data size, which helps alleviate the data sparsity problem. In addition, L1 regularization of the projection matrix is added to the objective function of each autoencoder layer in our model to enforce the sparstiy of projection matrix, and this in turn facilitates the discovery of the most relevant terms for reconstructing bullying terms. The main contributions of our work can be summarized as follows:

* Our proposed Semantic-enhanced Marginalized S- tacked Denoising Autoencoder is able to learn ro- bust features from BoW representation in an effi- cient and effective way. These robust features are learned by reconstructing original input from cor- rupted (i.e., missing) ones. The new feature space can improve the performance of cyberbullying de- tection even with a small labeled training corpus.
* Semantic information is incorporated into the re- construction process via the designing of semantic dropout noises and imposing sparsity constraints on mapping matrix. In our framework, high-quality semantic information, i.e., bullying words, can be extracted automatically through word embeddings. Finally, these specialized modifications make the new feature space more discriminative and this in turn facilitates bullying detection.
* Comprehensive experiments on real-data sets have verified the performance of our proposed model.

This paper is organized as follows. In Section 2, some re- lated work is introduced. The proposed Semantic-enhanced Marginalized Stacked Denoising Auto-encoder for cyber- bullying detection is presented in Section 3. In Section 4, experimental results on several collections of cyberbullying data are illustrated. Finally, concluding remarks are provid- ed in Section 5.

2 Related Work

This work aims to learn a robust and discriminative text rep- resentation for cyberbullying detection. Text representation and automatic cyberbullying detection are both related to our work. In the following, we briefly review the previous work in these two areas.

Text Representation Learning

In text mining, information retrieval and natural language processing, effective numerical representation of linguistic units is a key issue. The Bag-of-words (BoW) model is the most classical text representation and the cornerstone of some states- of-arts models including Latent Semantic Analysis (LSA) [18] and topic models [19], [20]. BoW model represents a document in a textual corpus using a vector of real numbers indicating the occurrence of words in the document. Although BoW model has proven to be efficient

and effective, the representation is often very sparse. To address this problem, LSA applies Singular Value Decompo- sition (SVD) on the word-document matrix for BoW model to derive a low-rank approximation. Each new feature is a linear combination of all original features to alleviate the sparsity problem. Topic models, including Probabilistic La- tent Semantic Analysis [21] and Latent Dirichlet Allocation [20], are also proposed. The basic idea behind topic models is that word choice in a document will be influenced by the topic of the document probabilistically. Topic models try to define the generation process of each word occurred in a document.

Similar to the approaches aforementioned, our proposed approach takes the BoW representation as the input. How- ever, our approach has some distinct merits. Firstly, the mul- ti-layers and non-linearity of our model can ensure a deep learning architecture for text representation, which has been proven to be effective for learning high-level features [22]. Second, the applied dropout noise can make the learned representation more robust. Third, specific to cyberbullying detection, our method employs the semantic information, including bullying words and sparsity constraint imposed on mapping matrix in each layer and this will in turn pro- duce more discriminative representation.

Cyberbullying Detection

With the increasing popularity of social media in recent years, cyberbullying has emerged as a serious problem afflicting children and young adults. Previous studies of cyberbullying focused on extensive surveys and its psy- chological effects on victims, and were mainly conducted by social scientists and psychologists [6], [23], [24], [25]. Although these efforts facilitate our understanding for cyberbullying, the psychological science approach based on personal surveys is very time-consuming and may not be suitable for automatic detection of cyberbullying. Since machine learning is gaining increased popularity in recent years, the computational study of cyberbullying has at- tracted the interest of researchers. Several research areas including topic detection and affective analysis are closely related to cyberbullying detection. Owing to their efforts, automatic cyberbullying detection is becoming possible. In machine learning-based cyberbullying detection, there are two issues: 1) text representation learning to transform each post/message into a numerical vector and 2) classifier train- ing. Xu et.al presented several off-the-shelf NLP solutions including BoW models, LSA and LDA for representation learning to capture bullying signals in social media [8]. As an introductory work, they did not develop specialized models for cyberbullying detection. Yin et.al proposed to combine BoW features, sentiment feature and contextual features to train a classifier for detecting possible harassing posts [10]. The introduction of the sentiment and contextual features has been proven to be effective. Dinakar et.al used Linear Discriminative Analysis to learn label specific features and combine them with BoW features to train a classifier [11]. The performance of label-specific features largely depends on the size of training corpus. In addition, they need to construct a bullyspace knowledge base to boost the performance of natural language processing methods. Although the incorporation of knowledge base can achieve a performance improvement, the construction of a complete and general one is labor-consuming. Nahar et.al proposed to scale bullying words by a factor of two in the original BoW features [12]. The motivation behind this work is quit similar to that of our model to enhance bullying features. However, the scaling operation in [12] is quite arbitrary. Ptaszynski et.al searched sophisticated patterns in a brute-force way [26]. The weights for each extracted pattern need to be calcu- lated based on annotated training corpus, and thus the per- formance may not be guaranteed if the training corpus has a limited size. Besides content-based information, Maral et.al also employ users' information, such as gender and history messages, and context information as extra features [13], [14]. Huang et.al also considered social network features to learn the features for cyberbullying detection [9]. The shared deficiency among these forementioned approaches is constructed text features are still from BoW representation, which has been criticized for its inherent over-sparsity and failure to capture semantic structure [18], [19], [20]. Differ- ent from these approaches, our proposed model can learn robust features by reconstructing the original data from corrupted data and introduce semantic corruption noise and sparsity mapping matrix to explore the feature structure which are predictive of the existence of bullying so that the learned representation can be discriminative.

3 Semantic-Enhanced Marginalized Stacked Denoising Auto-encoder

We first introduce notations used in our paper. Let $D = \{w_1, \ldots, w_d\}$ be the dictionary covering all the words exist- ing in the text corpus. We represent each message using a BoW vector $\mathbf{x} \in \mathrm{R}^d$. Then, the whole corpus can be denoted as a matrix: $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathrm{R}^{d \times n}$, where n is the number of available posts.

We next briefly review the marginalized stacked de- noising auto-encoder and present our proposed Semantic- enhanced Marginalized Stacked Denoising Auto-Encoder.

Marginalized Stacked Denoising Auto-encoder

Chen et.al proposed a modified version of Stacked Denois- ing Auto-encoder that employs a linear instead of a non- linear projection so as to obtain a closed-form solution [17]. The basic idea behind denoising auto-encoder is to recon- struct the original input from a corrupted one $\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_n$ with the goal of obtaining robust representation.

Marginalized Denoising Auto-encoder: In this mod- el, denoising auto-encoder attempts to reconstruct original data using the corrupted data via a linear projection. The projection matrix can be learned as:

$$\mathbf{W} = \underset{\mathbf{W}}{\operatorname{argmin}} \frac{1}{2n} \sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{W}\tilde{\mathbf{x}}_i\|_2 \quad (1)$$

where $\mathbf{W} \in \mathrm{R}^{d \times d}$. For simplicity, we can write Eq. (1) in matrix form as:

$$\mathbf{W} = \underset{\mathbf{W}}{\operatorname{argmin}} \frac{1}{2n} tr\left((\mathbf{X} - \mathbf{W}\tilde{\mathbf{X}})^{\mathrm{T}}(\mathbf{X} - \mathbf{W}\tilde{\mathbf{X}})\right) \quad (2)$$

where $\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_n]$ is the corrupted version of $\mathbf{X}$. It is easily shown that Eq. (2) is an ordinary least square problem having a closed-form solution:

$$\mathbf{W} = \mathbf{P}\mathbf{Q}^{-1} \tag{3}$$

where $\mathbf{P} = \mathbf{X}\tilde{\mathbf{X}}^{\mathrm{T}}$ and $\mathbf{Q} = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^{\mathrm{T}}$. In fact, this corruption can be marginalized over the noise distribution [17]. The more corruptions we take in the denoising auto-encoder, the more robust transformation can be learned. Therefore, the best choice is using infinite versions of corrupted data. If the data corpus is corrupted infinite times, the matrix P and Q are converged to their corresponding expectation, andEq.
(3) can be formulated as:

$$\mathbf{W} = E[\mathbf{P}]\,E[\mathbf{Q}]^{-1} \tag{4}$$

where $E[\mathbf{P}] = \sum_{n} E\,\mathbf{x}_i\tilde{\mathbf{x}}^{\mathrm{T}}$ and $E[\mathbf{Q}] = \sum_{n=1} E\,\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}^{\mathrm{T}}$. These expected matrices can be computed based on noise distribution. In [17], dropout noise is adopted to corrupt data samples by setting a feature to zero with a probability p. Assuming the scatter matrix of the original data samples is denoted as $\mathbf{S} = \mathbf{X}\mathbf{X}^{\mathrm{T}}$, the expected matrices can be computed as:

$$E[\mathbf{Q}]_{i,j} = \begin{cases} (1-p)^2\mathbf{S}_{i,j} & \text{if } i \neq j, \\ (1-p)\mathbf{S}_{i,j} & \text{if } i = j. \end{cases} \tag{5}$$

and

$$E[\mathbf{P}]_{i,j} = (1-p)\mathbf{S}_{i,j} \tag{6}$$

where i and j denotes the indices of features. It can be seen that it is very efficient to compute $\mathbf{W}$ by marginalizing dropout noise in denoising auto-encoder. After the mapping weights $\mathbf{W}$ are computed, a nonlinear squashing function, such as a hyperbolic tangent function, can be applied to de- rive the output of the marginalized denoisingauto-encoder:

$$\mathbf{H} = \tanh(\mathbf{W}\mathbf{X}) \tag{7}$$

Stacking Structure: Chen et.al [17] also proposed to apply stacking structures on marginalized denoising au-
toencoder, in which the output of the $(k-1)^{th}$ layer is fed as the input into the k^{th} layer. If we define the output of the k^{th} mDA as $\mathbf{H}_k$ and the original input as $\mathbf{H}_0$ respectively, the mapping between two consecutive layers is given as:

$$\mathbf{H}_k = \tanh(\mathbf{W}_k\mathbf{H}_{k-1}) \tag{8}$$

where $\mathbf{W}_k$ denotes the mapping in k^{th} layer. The model training can be done greedily layer by layer. This means that the mapping weights $\mathbf{W}_k$ is learned in a closed-form to
reconstruct the output of $(k-1)^{th}$mDA layer from its marginalized corruptions, as shown in Eq. (4). If the number of layers is set to L, the final representation for input data $\mathbf{X}$ is the concatenation of the uncorrupted original input and outputs of all layers as follows:

$$\mathbf{Z} = \begin{bmatrix} \mathbf{X} \\ \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_L \end{bmatrix} \tag{9}$$

where $\mathbf{Z} \in \mathrm{R}^{d(L+1)\times n}$. Each column of $\mathbf{Z}$ represents the final representation of each individual data sample.

Semantic Enhancement for mSDA

The advantage of corrupting the original input in mSDA can be explained by feature co-occurrence statistics. The co- occurrence information is able to derive a robust feature rep- resentation under an unsupervised learning framework, and this also motivates other state-of-the-art text feature learning methods such as Latent Semantic Analysis and topicmodels [18], [20]. As shown in Figure 1. (a), a denoising auto- encoder is trained to reconstruct these removed features values from the rest uncorrupted ones. Thus, the learned mapping matrix W is able to capture correlation between these removed features and other features. It is shown that
the learned representation is robust and can be regarded as a high level concept feature since the correlation informa-
tion is invariant to domain-specific vocabularies. We next describe how to extend mSDA for cyberbullying detection. The major modifications include semantic droupout noise and sparse mapping constraints.

Semantic Dropout Noise

The dropout noise adopted in mSDA is an uniform distribution, where each feature has the same probability to be removed. In cyberbullying detection, most bullying posts contain bullying words such as profanity words and foul languages. These bullying words are very predictive of the existence of cyberbullying. However, a direct use of these bullying features may not achieve good performance because these words only account for a small portion of the whole vocabulary and these vulgar words are only onekind of discriminative features for bullying [10], [26]. In other way, we can explore these cyberbullying words by using a different dropout noise that features corresponding to bullying words have a larger probability of corruption than other features. The imposed large probability on bullying words emphasizes the correlation between bullyingfeatures and normal ones. This kind of dropout noise can be denoted as semantic dropout noise, because semantic information is used to design dropout structure.

As shown in Figure 1. (b), the correlation between fea- tures can enable other normal words to predict bullying labels. Considering a simple but intuitive example, "Leave him alone, he is just a chink"[1], which is obviously a bullying message. However, the classifier will set the weight of the discriminative word "chink" to zero, if the small sized training corpus does not cover it. Our proposed smSDA can deal with the problem by learning a robust feature represen- tation, which is a high level concept representation. In the learned representation, the word "chink" are reconstruct- ed by context words co-occurring with the specific word ("chink") and the context words may be shared by other bullying words contained in training corpus. Therefore, the correlation explored by this auto-encoder structure enables the subsequent classifier to learn the discriminative word and improve the classification performance. In addition, the semantic dropout noise exploits the correlation between

1. *"Chink (also chinki, chinky, chinkie) is an English ethnic slur usually referring to a person of Chinese or East Asian ethnicity"* from Wikipedia

bullying features and normal features better and hence, facilitates cyberbullying detection.

Due to the introduced semantic dropout noise, the ex- pected matrices: $E[\mathbf{P}]$ and $E[\mathbf{Q}]$ will be computed slightly different from Eqs. (5) and (6). Assuming we have an available bullying words list and the correspondingfeatures set Z_b, the semantic dropout noise can be described as the following probability density function (PDF):

$$PDF = \begin{cases} p(\tilde{x}_d = 0) = p_n & \text{if } d \notin Z_b, \\ p(\tilde{x}_d = x_d) = 1 - p_n & \text{if } d \notin Z_b, \\ p(\tilde{x}_d = 0) = p_b & \text{if } d \in Z, \\ p(\tilde{x}_d = x_d) = 1 - p_b & \text{if } d \in Z_b, \end{cases} \quad (10)$$

where d denotes the feature set. Then these two marginal- ized matrices can be computed as:

(16)

where λ is a regularization parameter that controls the

$$E[\mathbf{Q}]_{i,j} = \begin{cases} (1-p_n)\mathbf{S}_{i,j} & \text{if } i = j \;\&\; i \notin Z_b, \\ (1-p_b)^2\mathbf{S}_{i,j} & \text{if } i \neq j \;\&\; \{i,j\} \cap Z_b = \emptyset, \\ (1-p_b)(1-p_n)\mathbf{S}_{i,j} & \text{if } \{i,j\} \notin Z_b \;\&\; \{i,j\} \cap Z_b \neq \emptyset, \\ (1-p_b)^2\mathbf{S}_{i,j} & \text{if } i \neq j \;\&\; \{i,j\} \in Z_b, \\ (1-p_b)\mathbf{S}_{i,j} & \text{if } i = j \;\&\; i \in Z_b \end{cases} \quad (11)$$

and

$$E[\mathbf{P}]_{i,j} = \begin{cases} (1-p_n)\mathbf{S}_{i,j} & \text{if } j \cap Z_b = \emptyset, \\ (1-p_b)\mathbf{S}_{i,j} & \text{if } j \cap Z_b \neq \emptyset. \end{cases} \quad (12)$$

where p_b and p_n are the probabilities of bullying features and normal features to be set to zero respectively, and $p_b > p_n$. Here, p_b and p_n are both tunable hyperparameters for our proposed smSDA.

Unbiased Semantic Dropout Noise As shown in Eq . (10), the corrupted data is biased, i.e., $E[\mathbf{X}] \neq E[\tilde{\mathbf{X}}]$. Here, we modified Eq. (10) to achieve an unbiased noise as follows:

$$PDF^{unbiased} = \begin{cases} p(\tilde{x}_d = 0) = p_n & \text{if } d \notin Z_b, \\ p(\tilde{x}_d = \frac{x_d}{1-p_n}) = 1 - p_n & \text{if } d \notin Z_b, \\ p(\tilde{x}_d = 0) = p_b & \text{if } d \in Z_b, \\ p(\tilde{x}_d = \frac{x_d}{1-p_b}) = 1 - p_b & \text{if } d \in Z_b, \end{cases} \quad (13)$$

It can be easily shown that under such a noise distribution, the corrupted data is unbiased now. These two marginalized matrices are re-formulated as:

$$E[\mathbf{Q}]^{unbiased}_{i,j} = \begin{cases} \frac{1}{1-p_n}\mathbf{S}_{i,j} & \text{if } i = j \;\&\; i \notin Z, \\ \frac{1}{1-p_b}\mathbf{S}_{i,j} & \text{if } i = j \;\&\; i \in Z_b, \\ \mathbf{S}_{i,j} & \text{if } i \neq j. \end{cases} \quad (14)$$

and

$$E[\mathbf{P}]^{unbiased}_{i,j} = \mathbf{S}_{i,j} \quad (15)$$

These two computed matrices will then be used to learn the mapping in each layer in our proposed smSDA.

Sparsity Constraints

In mSDA, the mapping matrix $\mathbf{W}$ is learned to reconstruc- t removed features from other uncorrupted features and hence is able to capture the feature correlation information. Here, we inject the sparsity constraints on the mapping weights $\mathbf{W}$ so that each row has a small number of nonzero elements. This sparsity constraint is quite intuitive because one word is only related to a small portion of vocabulary instead of the whole vocabulary. In our proposed smSDA, the sparsity constraint is realized by the incorporation of L1 regularization term into the objective function as in the lasso problem [27]. The optimization function for each layer in smSDA is given as follows:

$$\mathbf{W} = \arg\min_{w} \frac{1}{2n} tr\,(\mathbf{X} - \mathbf{W}\tilde{\mathbf{X}})^T(\mathbf{X} - \mathbf{W}\tilde{\mathbf{X}}) + \lambda\|\mathbf{W}\|_1$$

sparsity of $\mathbf{W}$. The larger the λ is, the sparser the mapping matrix $\mathbf{W}$ is. The solution to Eq. (16) is a very mature math problem: sparse least squares optimization, which has several effective and efficient computation methods [28], [29], [30]. Here, we adopt a method called Iterated Ridge Regression, which has been proven to be very efficient [30]. The method firstly introduces an approximation:

$$\| \approx \frac{\mathbf{w}_i^T \mathbf{w}_i}{\|\mathbf{w}_i\|_1} \quad (17)$$

where $\mathbf{w}_i$ denotes the i-th row in the whole matrix $\mathbf{W}$. By substituting this approximation Eq. (17) into the objective function Eq. (16), we yield an formulation similar to a Ridge Regression Problem [31], and the iteration steps to solve $\mathbf{W}$ is given as:

$$\mathbf{W}_k = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^T \left(\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T + \lambda\,\mathrm{diag}(|\mathbf{W}_{k-1}|)\right)^{-1} \quad (18)$$

where diag denotes the diagonal elements of a matrix, $\mathbf{W}_k$ and $\mathbf{W}_{k-1}$ denote the current step and the previous step estimations for mapping matrix $\mathbf{W}$ respectively. It is clear that the Eq. (18) can be easily formulated when the noise distribution is marginalized. Similar to Eq. (4), Eq. (18) can be written as:

$$\mathbf{W}_k = E[\mathbf{P}]\left(E[\mathbf{Q}] + \lambda\,\mathrm{diag}(|\mathbf{W}_{k-1}|)^{-1}\right)^{-1} \quad (19)$$

To speed up the convergence process, the initialization for $\mathbf{W}$ can be set to the L2 penalized solution for Eq. (2) as follows:

$$\mathbf{W}_0 = E[\mathbf{P}]\left(E[\mathbf{Q}] + \lambda\mathbf{I}\right)^{-1} \quad (20)$$

where $\mathbf{I}$ is an identify matrix. It can be shown that this itera- tion procedure can also marginalize the noise distribution easily, which can ensure an efficient and stable mapping learning.

ISBN 978-93-88808-61-3

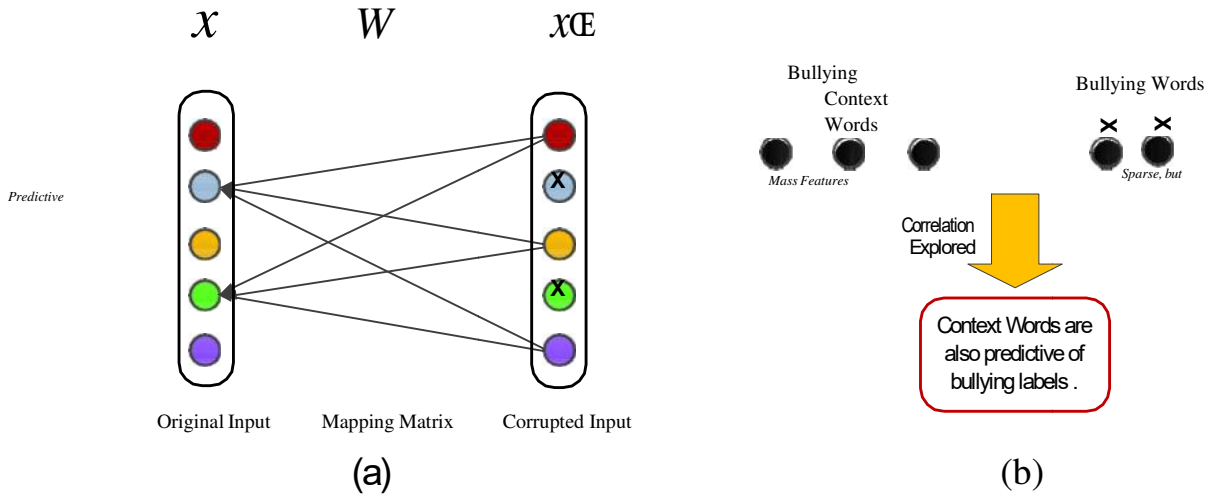


Fig. 1. Illustration of Motivations behind smSDA. In Figure 1(a), the cross symbol denotes that its corresponding feature is corrupted, i.e., turned off.

Construction of Bullying Feature Set

As analyzed above, the bullying features play an important role and should be chosen properly. In the following, the steps for constructing bullying feature set Z_b are given, in which the first layer and the other layers are addressed separately. For the first layer, expert knowledge and word embeddings are used. For the other layers, discriminative feature selection is conducted.

Layer One: firstly, we build a list of words with negative affective, including swear words and dirty words. Then, we compare the word list with the BoW features of our own corpus, and regard the intersections as bullying features. However, it is possible that expert knowledge is limited and does not reflect the current usage and style of cyberlanguage. Therefore, we expand the list of pre-defined insulting words, i.e. *insulting seeds*, based on word embeddingsas follows:

Word embeddings use real-valued and low-dimensional vectors to represent semantics of words [32], [33] . The well-trained word embeddings lie in a vector space where similar words are placed close to each other. In addition, the cosine similarity between word embeddings is able to quantify the semantic similarity between words. Considering the Interent messages are our interested corpus, we utilize a well-trained word2vec model on a large-scale twittercorpus containing 400 million tweets [34]. A visualization of some word embeddings after dimensionality reduction (PCA) is shown in Figure 2. It is observed that curse words form distinct clusters, which are also far away from normal words. Even insulting words are located at different regions due to different word usages and insulting expressions. In addition, since the word embeddings adopted here are trained in a large scale corpus from Twitter, the similarity captured by word embeddings can represent the specific language pattern. For example, the embedding of the misspelled word *fck* is close to the embedding of *fuck* so that the word *fck* can be automatically extracted based on word embeddings.

We extend the pre-defined *insulting seeds* based onword embeddings. For each insulting seed, similar words are extracted if their cosine similarities with insult seed exceed a predefined threshold. For bigram $w_l w_r$, we simply use an additive model to derive the corresponding embedding as follows:

$$\mathbf{v}(w_l w_r) = \mathbf{v}(w_l) + \mathbf{v}(w_r) \quad (21)$$

Finally, the constructed bullying features are used to train the first layer in our proposed smSDA. It includes two parts: one is the original *insulting seeds* based on domain knowledge and the other is the extended bullying words via word embeddings. The length of Z_b is k.

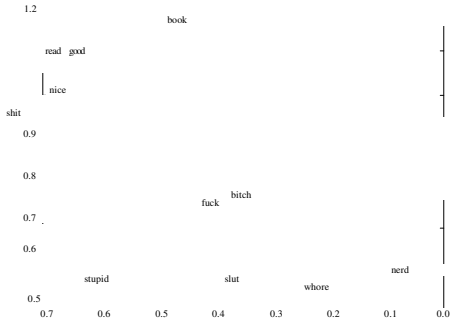


Fig. 2. Two dimensional visualization of our used word embeddings via PCA. Displayed terms include both bullying ones and normal ones. It shows that similar words are nearby vectors.

Subsequent Layers: we perform feature selection using Fisher score to select ''bullying'' features. Fisher score is an univariate metric reflecting the discriminative power of a feature [35], [36]. For the r^{th} feature, the corresponding Fisher score can be computed based on training data with

$$F_r = \frac{\sum_{i=1}^{c} n_i (\mu_{ir} - \mu_r)^2}{\sum_{i=1}^{c} n_i \sigma_{ir}^2} \quad (22)$$

where c denotes the number of classes and n_i represent the number of data in class i. μ and μ_i denote the mean of entire data and class i for the r^{th} feature, and σ_i is the variance of class i on r^{th} feature. After Fisher scores are estimated, fea- tures with top k scores are selected as ''bullying'' features, where ''bullying'' is generalized as discriminative.

smSDA for Cyberbullying Detection

In section 3.3, we propose the Semantic-enhanced Marginal- ized Stacked Denoising Auto-encoder (smSDA). In this subsection, we describe how to leverage it for cyberbully- ing detection. smSDA provides robust and discriminative representations The learned numerical representations can then be fed into Support Vector Machine (SVM). In the new space, due to the captured feature correlation and semantic information, the SVM, even trained in a small size of training corpus, is able to achieve a good performance on testing documents(this will be verified in the following experiments). The detailed steps of our model are provided below:

Assuming the first n_l posts are labeled and the corre- sponding vector of binary labels is $\mathbf{y} = y_1, \ldots, y_{n_l}$. The binarylabel 1 or 0 indicates the postis oris not acyberbully- ing one. Here, n_l n, which means the labeled posts have a small size. The bullying feature set Z_b is constructed in a layer-wise way. Based on prior knowledge, we construct are selected as *insulting seeds*. The insulting seeds are then expanded and refined automatically via word embeddings, which defines the bullying features Z_b for layer one. The experiments in Section 4 will show that the construction of the set Z_b is very simple and efficient with litter human labor. For the subsequent layers, after obtaining the output of each layer, the set Z_b is updated using feature ranking with Fish score according to Eq. (22).

Based on predefined dropout probabilities for bullying features and other normal features p_b and p_n and the bullying feature set Z_b, we compute these two expected matrices $E[\mathbf{P}]$ and $E[\mathbf{Q}]$ according to Eqs. (12) and (11), if the semantic dropout noise is adopted. When it comes to the unbiased semantic dropout noise, Eqs. (14) and (15) instead of Eqs. (12) and (11) are used to compute these two expected matrices. Then, we iteratively perform Eq. (21) for T_{max} times, where the initial value for $\mathbf{W}$ is calculated based on Eq. (20). When the mapping matrix is learned, the output of each layer is given according to Eq. (8). Due to the stacking structure, the output of L layers and the initial input are concatenated together to form the final representation $\mathbf{Z} \in R^{d(L+1)\times n}$ following Eq. (9). It is clear that the new space has a dimension of $(L + 1)d$. A linear SVM [37] is trained on the training corpus, i.e. the first n_l columns in $\mathbf{Z}$ and tested on the rest data samples.

Merits of smSDA

Some important merits of our proposed approach are sum- marized as follows:

1) Most cyberbullying detection methods rely on the BoW model. Due to the sparsity problems of both data and features, the classifier may not be trained very well. Stacked densoing autoencoder (SDA), as an unsupervised representation learning method, is able to learn a robust feature space. In SDA, the feature correlation is explored by the reconstruction of corrupted data. The learned robust feature rep- resentation can then boost the training of classifier and finally improve the classification accuracy. In addition, the corruption of data in SDA actually generates artificial data to expand data size, which alleviate the small size problem of training data.
2) For cyberbullying problem, we design semantic dropout noise to emphasize bullying features in the new feature space, and the yielded new represen- tation is thus more discriminative for cyberbullying detection.
3) The sparsity constraint is injected into the solution of mapping matrix $\mathbf{W}$ for each layer, considering each word is only correlated to a small portion of the whole vocabulary. We formulate the solution for the mapping weights $\mathbf{W}$ as an Iterated Ridge Regres- sion problem, in which the semantic dropout noise distribution can be easily marginalized to ensurethe efficient training of our proposed smSDA.
4) Based on word embeddings, bullying features can be extracted automatically. In addition, the possible limitation of expert knowledge can be alleviated by the use of word embedding.

4 EXPERIMENTS

In this section, we evaluate our proposed semantic- enhanced marginalized stacked denoising auto-encoder (smSDA) with two public real-world cyberbullying corpora. We start by describing the adopted corpora and experimen- tal setup. Experimental results are then compared with other baseline methods to test the performance of our approach. At last, we provide a detailed analysis to explain the good performance of our method.

Descriptions of Datasets

Two datasets are used here. One is from *Twitter* and another is from *MySpace* groups. The details of these two datasets are described below:

***Twitter* Dataset**: *Twitter* is ''a real-time information net- work that connects you to the latest stories, ideas, opin- ions and news about what you find interesting'' (https: //about.twitter.com/). Registered users can read and post tweets, which are defined as the messages posted on *Twitter* with a maximum length of 140 characters.

The *Twitter* dataset is composed of tweets crawled by the public *Twitter* stream API through two steps. In Step 1, keywords starting with ″bull″ including ″bully″, ″bullied″ and ″bullying″ are used as queries in Twitter to preselect some tweets that potentially contain bullying contents. Re- tweets are removed by excluding tweets containing the acronym ''RT''. In Step 2, the selected tweets are manually labeled as bullying trace or non-bullying trace based on the contents of the tweets. 7321 tweets are randomly sampled from the whole tweets collections from August 6, 2011 to

ISBN 978-93-88808-61-3

August 31, 2011 and manually labeled[2]. It should be pointed out here that labeling is based on bullying traces. A bullying trace is defined as the response of participants to their bullying experience. Bullying traces include not only mes- sages about direct bullying attack, but also messages about reporting a bullying experience, revealing self as a victim et. al. Therefore, bullying traces far exceed the incidents of cyberbullying. Automatic detection of bullying traces are valuable for cyberbullying research [38]. Some examples of bullying traces are shown in Figure 3. To preprocess these tweets, a tokenizer is applied without any stemming or stopword removal operations. In addition, some special characters including user mentions, URLS and so on are replaced by predefined characters, respectively. The features are composed of unigrams and bigrams that should appear at least twice and the details of preprocessing can be found in [8]. The statistics of this dataset can be found in Table 1. ***MySpace* Dataset**: *MySpace* is another web2.0 social net- working website. The registered accounts are allowed to view pictures, read chat and check other peoples' profile information.

The *MySpace* dataset is crawled from *MySpace* groups. Each group consists of several posts by different users, which can be regarded as a conversation about one topic. Due to the interactive nature behind cyberbullying, each data sample is defined as a window of 10 consecutive posts and the windows are moved one post by one post so that we got multiple windows [39]. Then, three people labeled the data for the existence of bullying content independently. To be objective, an instance is labeled as cyberbullying only if at least 2 out of 3 coders identify bullying content in the windows of posts. The raw text for these data, as XML files, have been kindly provided by Kontostathis et.al[3]. The XML files contain information about the posts, such as post text, post data, and users' information, which are put into 11 packets. Some posts in *MySpace* are shown in Figure 4.

Here, we focus on content-based mining, and hence, we only extract and preprocess the posts' text. The preprocessing steps of the *MySpace* raw text includetokenization, dele- tion of punctuation and special characters. The unigrams and bigrams features are adopted here. The threshold for negligible low-frequency terms is set to 20, considering one post occurred in a long conversation will occur in at least ten windows. The details of this dataset is shown in Table 1. Since there were no standard splits of training vs. test datasets in our adopted *Twitter* and *MySpace* corpora, we need to define the training and testing datasets. As analyzed above that the lack of labeled training corpus hinders the development of automatic cyberbullying detection, the sizes of training corpus are all controlled to be very small in our experiments. For *Twitter* dataset, we randomly select 800 instances, which accounts for 12% of the whole corpus, as the training data and the rest data samples are used as testing data. To reduce variance, the process is repeated ten times so that we can have ten *sub*-datasets from *Twitter* data. For *MySpace* dataset, we also randomly pick 400 data samples as the training corpus and use the rest data for testing. The process is repeated ten times to generate ten *sub*-datasets constructed from *MySpace* data. Finally, we have twenty *sub*-datasets, in which ten datasets are from *Twitter* corpus and another ten datasets are from *MySpace* corpus.

2. The dataset: **bullyingV3.0**, has been kindly provided at http:// research.cs.wisc.edu/bullying/data.html

3. The dataset: **MySpace Group**, has been kindly provided at http: //www.chatcoder.com/DataDownload

TABLE 1
Statistical Properties of the two datasets.

Statistics	*Twitter*	*MySpace*
Feature No.	4413	4240
Sample No.	7321	1539
Bullying Instances	2102	398

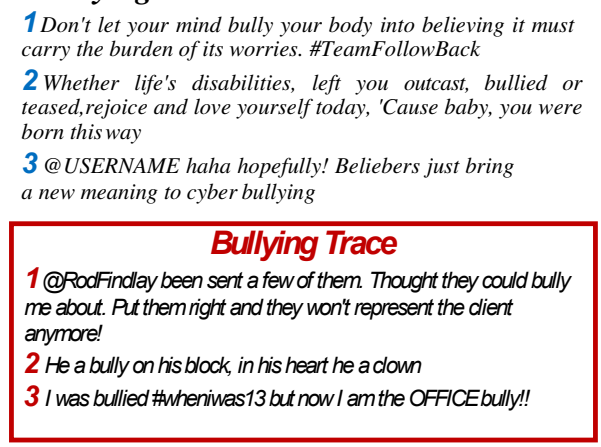


Fig. 3. Some Examples from *Twitter* Datasets. Three of them are non-bullying traces. And the other three are bullying traces.

Experimental Setup

Here, we experimentally evaluate our smSDA on two cyber-bullying detection corpora. The following methods will be compared.

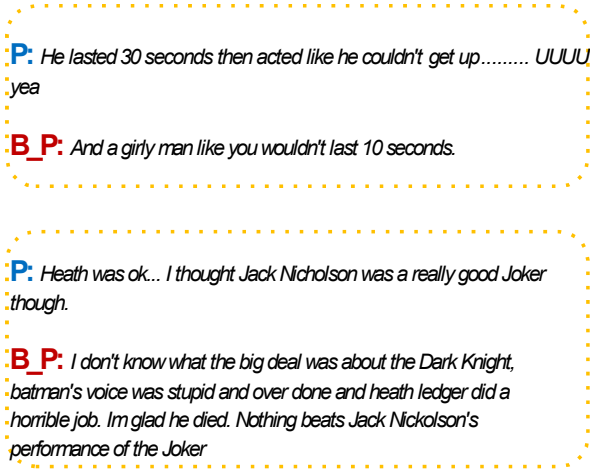


Fig. 4. Some Examples from *MySpace* Datasets. Two Conversions are Displayed and each one includes a normal post (P) and a bullyingpost ($B\,P$) .

* BWM: Bullying word matching. If the message con- tains at least one of our defined bullying words, it will be classified as bullying.
* BoW Model: the raw BoW features are directly fed into the classifier.
* Semantic-enhanced BoW Model: This approach is referred in [12]. Following the original setting, we scale the bullying features by a factor of 2.
* LSA: Latent Semantic Analysis [18].
* LDA: Latent Dirchilet Allocation [20]. Our implementation of LDA is based on *Gensim*[4].
* mSDA: marginalized stacked denoising autoen- coder [17].
* smSDA and smSDA$_u$: semantic-enhanced marginal- ized denoising autoencoder that utilizes semantic dropout noise and unbiased one, respectively.

For LSA and LDA, the number of latent topics are both set to 100. In LDA, we set hyperparameter α for document topic multinomial and hyperparameter η for word topic multinomial to 1 and 0.01, respectively. For mSDA[5], the noise intensity is set to 0.5 and the number of layers for *Tweets* and *MySpace* datasets are both set to 2. Here, the number of layers is only set to be a moderate number instead of a large one, considering a large final dimension will impose a computational burden on the subsequent classifier training.

For our proposed methods including smSDA and smSDA$_u$: the noise intensity and the number of layers are set to the same values as in mSDA to give a fair comparison. The bullying noise intensity is set to 0.8, which is larger than 0.5. The hyperparameters λ that controls the sparsity of the transformation matrix are set to 1 for all layers. The number of iteration step for solving lasso problems is set to
20. To construct the bullying features Z_b for the first layer,

whose word cloud visualization is shown in Figure 5. The intersections between BoW features of our own corpus and the predefined bullying word list are firstly obtained. Then, as described in 3.3, they are extended and refined based on word embeddings to form the final bullying features. The threshold for cosine similarity is set to 0.8. The word cloud

visualizations for the final bullying features in *Twitter* and *MySpace* datasets are shown in Figures 6 and 7, respective- ly. The bullying features used in Semantic-enhanced BoW Model are the same as those in smSDA.

Linear SVM [37] is then applied to the new feature space generated by the above mentioned approaches. In linear SVM, we search the best regularization parameter C from
$\{0.0001, 0.001, 0.01, 0.1, 1\}$. To evaluate the performance of these methods on binary classification, classification accuracy is employed. Considering both datasets have the class imbalance problem, we also introduce F1-Score, which is a balance between precision and recall, to evaluate the performance of all compared approaches.

4. https://radimrehurek.com/gensim/index.html
5. The code has been kindly provided athttp://research.cs.wisc.edu/bullying/data.html
6. A collection of insulting words can be found in the website: http://www.noswearing.com/dictionary

Fig. 5. Word Cloud Visualization of the List of Words with Negative Affective.

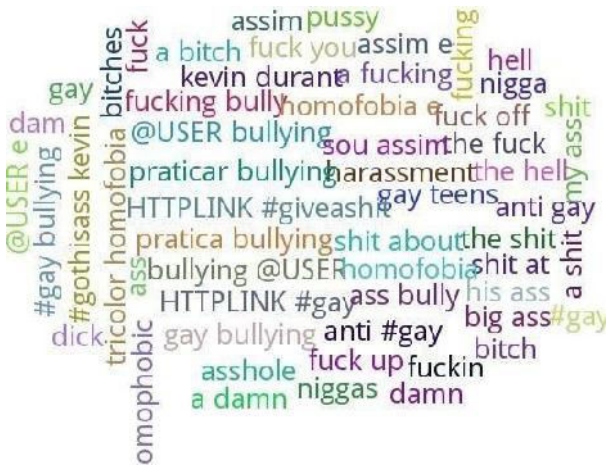


Fig. 6. Word Cloud Visualization of the Bullying Features in *Twitter* Datasets.

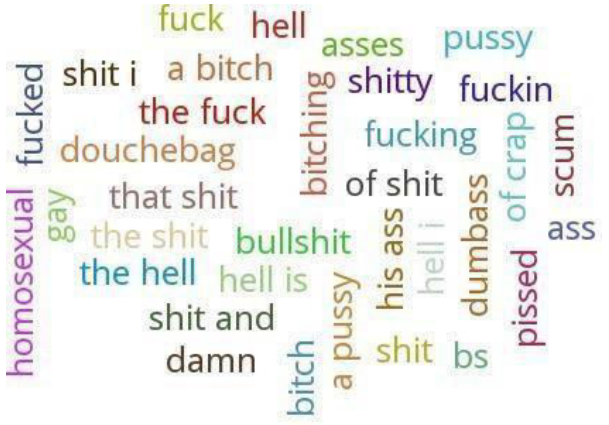


Fig. 7. Word Cloud Visualization of the Bullying Features in *MySpace* Datasets.

ISBN 978-93-88808-61-3

Experimental Results

In this section, we show a comparison of our proposed smSDA method with six benchmark approaches on *Twitter* and *MySpace* datasets. The average results, for these two datasets, on classification accuracy and F1 score are shown in Table 2. Figures 8 and 9 show the results of seven compared approaches on all *sub*-datasets constructed from *Twitter* and *MySpace* datasets, respectively. Since BWM does not require training documents, its results over the whole corpus are reported in Table 2. It is clear that our approaches outperform the other approaches in these two *Twitter* and *MySpace* corpora.

The first observation is that semantic BoW model (sBow) performs slightly better than BoW. Based on BoW, sBoWjust arbitrarily scale the bullying features by a factor of 2. This means that semantic information can boost theperformance of cyberbullying detection. For a fair comparison, the bul- lying features used in our method and sBoW are unified to be the same. Our approaches, especially smSDA, gains a significant performance improvement compared to sBoW. This is because bullying features only account for a small portion of all features used. It is difficult to learn robust features for small training data by intensifying each bullying features' amplitude. Our approach aims to find the cor- relation between normal features and bullying features by reconstructing corrupted data so as to yield robust features. In addition, Bullying Word Matching (BWM), as a simple and intuitive method of using semantic information, gives the worst performance. In BWM, the existence of bullying words are defined as rules for classification. It shows that only an elaborated utilization of such bullying words in- stead of a simple one can help cyberbullying detection.

We also compare our methods with two stat-of-arts text representation learning methods LSA and LDA. These two methods do not produce good performance on all datasets. This may be because that both methods belong to dimen- sionality reduction techniques, which are performed on the document-word occurrence matrix. Although the two meth- ods try to minimize the reconstruction error as our approach does, the optimization in LSA and LDA is conducted after dimensionality reduction. The reduced dimension is a key parameter to determine the quality of learned feature space. Here, we fix the dimension of latent space to 100. Therefore, a deliberate searching for this parameter which may im- prove the performances of LSA and LDA and the selection of hyperparameter itself is another tough research topic. Another reason may be that the data samples are small (less than 2000) and the length of each Internet message is short (For *Twitter*, maximum length is 140 characters),and thus the constructed document-word occurrence matrix may not represent the true co-occurrence of terms.

Deep learning methods including mSDA and smSDA generally outperform other standard approaches. This trend is particularly prominent in F1 measure because cyberbul- lying detection problems are class-imbalance. The larger improvements on F1 score verify the performance of our approach further. Deep learning models have achieved re- markable performance in various scenarios with its own robust feature learning ability [22]. mSDA is able to cap- ture the correlation between input features and combine

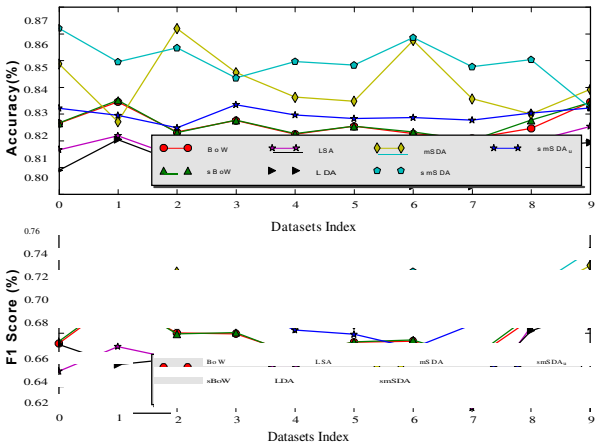


Fig. 8. Classification Accuracies and F1 Scores of All Compared Meth- ods on *Twitter* Datasets.

the correlated features by reconstructing masking feature values from uncorrupted feature values. Further, the stack- ing structure and the nonlinearity contribute to mSDA's ability for discovering complex factors behind data. Based on mSDA, our proposed smSDA utilizes semantic dropout noise and sparsity constraints on mapping matrix, in which the efficiency of training can be kept. This extension leads to a stable performance improvement on cyberbullying de- tection and the detailed analysis has been provided in the following section.

We compare the performances of smSDA and $smSDA_u$, which adopt biased semantic dropout noise and unbi- ased semantic dropout noise, respectively. The results have shown that $smSDA_u$ performs slightly worse than smSDA. This may be explained by the fact that the unbiased semantic dropout noise cancels the enhancement of bullying features. As shown in Eq. (14), the off-diagonal elements in the matrix $\mathbf{x}_i\tilde{\mathbf{x}}^T$ that are used to compute mapping weights are the same, which can not contribute to the reinforcement of bullying features.

Analysis of Semantic Extension

As shown in the section 4.3, the semantic extension can boost the performance on classification results for cyber- bullying detection. In this section, we discuss the advan- tages of this extension qualitatively. In our proposed smS- DA, because of the semantic dropout noise and sparsity constraints, the learned representation is able to discover the correlation between words containing latent bullying semantics. Table 3 shows the reconstruction terms of three example bullying words for mSDA and smSDA, respec- tively. In this example, one-hot vector is used as input, which represents a document containing one bullyingword. Table 3 lists the reconstructed terms in decreasing order of their feature values, which represents the strength of their correlations with the input word. The results are obtained using one layer architecture without non-linear activation considering the raw terms directly correspondto

ISBN 978-93-88808-61-3

TABLE 2
Accuracies (%), and F1 Scores (%) for Compared Methods on *Twitter* and *MySpace* Datasets. The Mean Values are Given, respectively. Bold Face Indicates Best Performance.

Dataset	Measures	BWM	BoW	sBow	LSA	LDA	mSDA	$smSDA_u$	smSDA
Twitter	Accuracies	69.3	82.6	82.7	81.6	81.1	84.1	82.9	**84.9**
	F1 Scores	16.1	68.1	68.3	65.8	66.1	70.4	69.3	**71.9**
MySpace	Accuracies	34.2	80.1	80.1	77.7	77.8	87.8	88.0	**89.7**
	F1 Scores	36.4	41.2	42.5	45.0	43.1	76.1	76.0	**77.6**

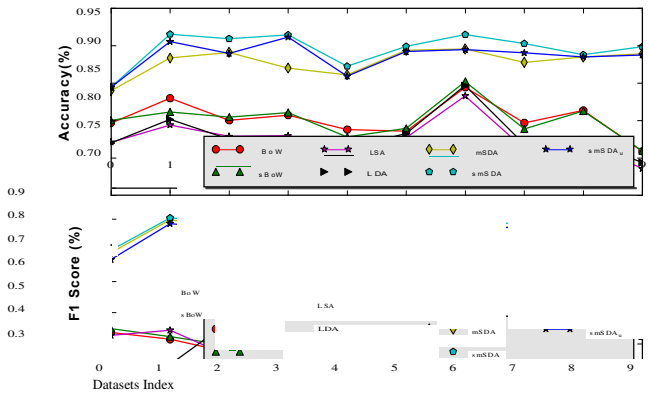


Fig. 9. Classification Accuracies and F1 Scores of All Compared Meth- ods on *MySpace* Datasets.

each output dimension under such a setting. It is shown that these reconstructed words discovered by smSDA are more correlated to bullying words than those by mSDA. For example, *fucking* is reconstructed by *because*, *friend*, *off*, *gets* in mSDA. Except *off*, the other three words seem to be unreasonable. However, in smSDA, *fucking* is reconstructed by *off*, *pissed*, *shit* and *of*. The occurrence of the term *of* may be due to the frequent misspelling in Internet writing. It is obvious that the correlation discovered by smSDA is more meaningful. This indicates that smSDA can learn the
words' correlations which may be the signs of bullying semantics, and therefore the learned robust features boost

the performance on cyberbullying detection.

5 Conclusion

This paper addresses the text-based cyberbullying detection problem, where robust and discriminative representations of messages are critical for an effective detection system.
By designing semantic dropout noise and enforcing spar- sity, we have developed semantic-enhanced marginalized denoising autoencoder as a specialized representation learn-
ing model for cyberbullying detection. In addition, word embeddings have been used to automatically expand and refine bullying word lists that is initialized by domain knowledge. The performance of our approaches has been experimentally verified through two cyberbullying corpora from social medias: *Twitter* and *MySpace*. As a next step we are planning to further improve the robustness of the

TABLE 3
Term Reconstruction on *Twitter* datasets. Each Row Shows Specific Bullying Word, along with Top-4 Reconstructed Words (ranked with their frequency values from top to bottom) via mSDA (left column) and smSDA (right column).

Bullying Words	Reconstructed Words for mSDA	Reconstructed Words for smSDA
bitch	*@USER shut friend tell*	*@USER HTTPLINK fuck up shut*
fucking	*because friend off gets*	*off pissed shit of*
shit	*some big with lol*	*abuse this shit shit lol big*

learned representation by considering word order in mes- sages.

Acknowledgments

We thank Junming Xu and Prof. Jerry Zhu for their *Twitter* datasets.

References

[1] A. M. Kaplan and M. Haenlein, "Users of the world, unite! the challenges and opportunities of social media," *Business horizons*, vol. 53, no. 1, pp. 59–68, 2010.
[2] R. M. Kowalski, G. W. Giumetti, A. N. Schroeder, and M. R. Lattanner, "Bullying in the digital age: A critical review and meta- analysis of cyberbullying research among youth." 2014.
[3] M. Ybarra, "Trends in technology-based sexual and non-sexual aggression over time and linkages to nontechnology aggression," *National Summit on Interpersonal Violence and Abuse Acrossthe Lifespan: Forging a Shared Agenda*, 2010.
[4] B. K. Biggs, J. M. Nelson, and M. L. Sampilo, "Peer relations in the anxiety–depression link: Test of a mediation model," *Anxiety, Stress, & Coping*, vol. 23, no. 4, pp. 431–447, 2010.
[5] S. R. Jimerson, S. M. Swearer, and D. L. Espelage, *Handbook of bullying in schools: An international perspective*. Routledge/Taylor & Francis Group, 2010.
[6] G. Gini and T. Pozzoli, "Association between bullying and psy- chosomatic problems: A meta-analysis," *Pediatrics*, vol. 123, no. 3, pp. 1059–1065, 2009.
[7] A. Kontostathis, L. Edwards, and A. Leatherman, "Text mining and cybercrime," *Text Mining: Applications and Theory. John Wiley & Sons, Ltd, Chichester, UK*, 2010.
[8] J.-M. Xu, K.-S. Jun, X. Zhu, and A. Bellmore, "Learning from bullying traces in social media," in *Proceedings of the 2012 conference of the North American chapter of the association for computational linguistics: Human language technologies*. Association for Compu- tational Linguistics, 2012, pp. 656–666.
[9] Q. Huang, V. K. Singh, and P. K. Atrey, "Cyber bullying detection using social and textual analysis," in *Proceedings of the 3rd Inter- national Workshop on Socially-Aware Multimedia*. ACM, 2014, pp. 3– 6.

[10] D. Yin, Z. Xue, L. Hong, B. D. Davison, A. Kontostathis,and L. Edwards, "Detection of harassment on web 2.0," *Proceedings of the Content Analysis in the WEB*, vol. 2, pp. 1–7, 2009.

[11] K. Dinakar, R. Reichart, and H. Lieberman, "Modeling the detec- tion of textual cyberbullying." in *The Social Mobile Web*, 2011.

[12] V. Nahar, X. Li, and C. Pang, "An effective approach for cy- berbullying detection," *Communications in Information Science and Management Engineering*, 2012.

[13] M. Dadvar, F. de Jong, R. Ordelman, and R. Trieschnigg, "Im- proved cyberbullying detection using gender information," in *Proceedings of the 12th -Dutch-Belgian Information Retrieval Workshop (DIR2012)*. Ghent, Belgium: ACM, 2012.

[14] M. Dadvar, D. Trieschnigg, R. Ordelman, and F. de Jong, "Im- proving cyberbullying detection with user context," in *Advances in Information Retrieval*. Springer, 2013, pp. 693–696.

[15] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *The Journal of Machine Learning Research*, vol. 11, pp. 3371–3408, 2010.

[16] P. Baldi, "Autoencoders, unsupervised learning, and deep archi- tectures," *Unsupervised and Transfer Learning Challenges in Machine Learning, Volume 7*, p. 43, 2012.

[17] M. Chen, Z. Xu, K. Weinberger, and F. Sha, "Marginalized de- noising autoencoders for domain adaptation," *arXiv preprint arX- iv:1206.4683*, 2012.

[18] T. K. Landauer, P. W. Foltz, and D. Laham, "An introduction to latent semantic analysis," *Discourse processes*, vol. 25, no. 2-3, pp. 259–284, 1998.

[19] T. L. Griffiths and M. Steyvers, "Finding scientific topics," *Proceed- ings of the National academy of Sciences of the United States of America*, vol. 101, no. Suppl 1, pp. 5228–5235, 2004.

[20] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *the Journal of machine Learning research*, vol. 3, pp. 993–1022, 2003.

[21] T. Hofmann, "Unsupervised learning by probabilistic latent se- mantic analysis," *Machine learning*, vol. 42, no. 1-2, pp. 177–196, 2001.

[22] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 8, pp. 1798–1828, 2013.

[23] B. L. McLaughlin, A. A. Braga, C. V. Petrie, M. H. Moore *et al.*, *Deadly Lessons:: Understanding Lethal School Violence*. National Academies Press, 2002.

[24] J. Juvonen and E. F. Gross, "Extending the school ground- s?bullying experiences in cyberspace," *Journal of School health*, vol. 78, no. 9, pp. 496–505, 2008.

[25] M. Fekkes, F. I. Pijpers, A. M. Fredriks, T. Vogels, and S. P. Verloove-Vanhorick, "Do bullied children get ill, or do ill chil- dren get bullied? a prospective cohort study on the relationship between bullying and health-related symptoms," *Pediatrics*, vol. 117, no. 5, pp. 1568–1574, 2006.

[26] M. Ptaszynski, F. Masui, Y. Kimura, R. Rzepka, and K. Araki, "Brute force works best against bullying," in *Proceedings of IJCAI 2015 Joint Workshop on Constraints and Preferences for Configuration and Recommendation and Intelligent Techniques for Web Personaliza- tion*. ACM, 2015.

[27] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[28] C. C. Paige and M. A. Saunders, "Lsqr: An algorithm for sparse linear equations and sparse least squares," *ACM Transactions on Mathematical Software (TOMS)*, vol. 8, no. 1, pp. 43–71, 1982.

[29] M. A. Saunders *et al.*, "Cholesky-based methods for sparse least squares: The benefits of regularization," *Linear and Nonlinear Con- jugate Gradient-Related Methods*, pp. 92–100, 1996.

[30] J. Fan and R. Li, "Variable selection via nonconcave penalized like- lihood and its oracle properties," *Journal of the American statistical Association*, vol. 96, no. 456, pp. 1348–1360, 2001.

[31] C. Vogel, *Computational Methods for Inverse Problems*. Society for Industrial and Applied Mathematics, 2002. [Online]. Available: http://epubs.siam.org/doi/abs/10.1137/1.9780898717570

[32] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient esti- mation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[33] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their com- positionality," in *Advances in neural information processing systems*, 2013, pp. 3111–3119.

[34] F. Godin, B. Vandersmissen, W. De Neve, and R. Van de Walle, "Named entity recognition for twitter microposts using distributed word representations," in *Proceedings of the Workshop on Noisy User-generated Text*. Beijing, China: Association for Computational Linguistics, July 2015, pp. 146–153. [Online]. Available: http://www.aclweb.org/anthology/W15-4322

[35] T. H. Dat and C. Guan, "Feature selection based on fisher ratio and mutual information analyses for robust brain computer interface," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 1. IEEE, 2007, pp. I–337.

[36] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.

[37] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[38] J. Sui, "Understanding and fighting bullying with machine learn- ing," Ph.D. dissertation, THE UNIVERSITY OF WISCONSIN- MADISON, 2015.

[39] J. Bayzick, A. Kontostathis, and L. Edwards, "Detecting the pres- ence of cyberbullying using computer software," in *Proceedings of the ACM WebSci'11*. Koblenz, Germany: ACM, June 2011, pp. 1–2.

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

Efficient and Expressive Keyword Search Over Encrypted Data in Cloud

L.Ananth Lakshmi
M.Tech.
Schholar, CSE Department,
Malla Reddy College of Engineering

Abstract—Searchable encryption allows a cloud server to conduct keyword search over encrypted data on behalf of the data users without learning the underlying plaintexts. However, most existing searchable encryption schemes only support single or conjunctive keyword search, while a few other schemes that are able to perform expressive keyword search are computationally inefficient since they are built from bilinear pairings over the composite-order groups. In this paper, we propose an expressive public-key searchable encryption scheme in the prime-order groups, which allows keyword search policies (i.e., predicates, access structures) to be expressed in conjunctive, disjunctive or any monotonic Boolean formulas and achieves significant performance improvement over existing schemes. We formally define its security, and prove that it is selectively secure in the standard model. Also, we implement the proposed scheme using a rapid prototyping tool called Charm [37], and conduct several experiments to evaluate it perform ance. The results demonstrate that our scheme is much more efficient than the ones built over the composite-order groups.

Index Terms—Searchable encryption, cloud computing, expressiveness, attribute-based encryption.

1 INTRODUCTION

Consider a cloud-based healthcare information system that hosts outsourced personal health records (PHRs) from various healthcare providers. The PHRs are encrypted in order to comply with privacy regulations like HIPAA. In order to facilitate data use and sharing, it is highly desir- able to have a searchable encryption (SE) scheme which allows the cloud service provider to search over encrypted PHRs on behalf of the authorized users (such as medical researchers or doctors) without learning information about the underlying plaintext. Note that the context we are con- sidering supports private data sharing among multiple data providers and multiple data users. Therefore, SE schemes in the private-key setting [1], [2], [3], which assume that a single user who searches and retrieves his/her own data, are not suitable. On the other hand, private information retrieval (PIR) protocols [4], [5], [6], which allow users to retrieve a certain data-item from a database which publicly stores data without revealing the data-item to the database administrator, are also not suitable, since they require the data to be publicly available. In order to tackle the keyword search problem in the cloud-based healthcare information system scenario, we resort to public-key encryption with keyword search (PEKS) schemes, which is firstly proposed in [7]. In a PEKS scheme, a ciphertext of the keywords called "PEKS ciphertext" is appended to an encrypted PHR. To retrieve all the encrypted PHRs containing a keyword, say "Diabetes", a user sends a "trapdoor" associated with a search query on the keyword "Diabetes" to the cloud service provider, which selects all the encrypted PHRs containing the keyword "Diabetes" and returns them to the user while without learning the underlying PHRs. However, the solu- tion in [7] as well as other existing PEKS schemes which improve on [7] only support equality queries [8].

Set intersection and meta keywords[1] [9], [10] can be used for conjunctive keyword search. However, the approach based on set intersection leaks extra information to the cloud server beyond the results of the conjunctive query, whilst the approach using meta keywords require 2^m meta keywords to accommodate all the possible conjunctive queries for m keywords. In order to address the above deficiencies in conjunctive keyword search, schemes such as the ones in [11], [12] were put forward in the public-key setting.

Ideally, in the practical applications, search predicates (i.e., policies) should be expressive such that they can be expressed as conjunction, disjunction or any Boolean for- mulas[2] of keywords. In the above cloud-based healthcare system, to find the relationship between diabetes and age or weight, a medical researcher may issue a search query with an access structure (i.e., predicate) ("Illness = Diabetes" AND ("Age = 30" OR "Weight = 150-200")). SE schemes supporting expressive keyword access structures were p- resented in [8], [13], [14], [15]. Unfortunately, the scheme in [13] has exponentially increasing complexity [16], while the schemes in [8], [14], [15] are based on the inefficient bilinear pairing over composite-order groups [17]. Though there exist techniques [17] to convert pairing-based schemes from composite-order groups to prime-order groups, there is still a significant performance degradation due to the

1. Meta keywords are composed of several keywords. For example, a that contains the document keywords "Bob", "urgent" and "finance" may be augmented with the meta-keyword "Bob: urgent: finance"

required size of the special vectors [18].

In this paper, we propose a public-key based expressive SE scheme in prime-order groups, which is especially suit- able for keyword search over encrypted data in scenarios of multiple data owners and multiple data users such as the cloud-based healthcare information system that hosts outsourced PHRs from various healthcare providers.

Overview of Our Proposed Scheme

Our expressive SE scheme consists of a trusted trapdoor generation center which publishes a public system parame- ter and keeps a master key in secret, a cloud server which stores and searches encrypted data on behalf of data users, multiple data owners who upload encrypted data to the cloud, and multiple data users who would like to retreive encrypted data containing certain keywords. To outsource an encrypted document to the cloud, a data owner appends the encrypted document with keywords encrypted under the public parameter and uploads the combined encrypted document and encrypted keywords to the cloud. To retrieve all the encrypted documents containing keywords satisfying a certain access structure (i.e., predicate or policy) such as ("Illness = Diabetes" AND ("Age = 30" OR "Weight = 150-200")), a data user first obtains a trapdoor associated with the access structure from the trapdoor generation center and then sends the trapdoor to the cloud server. The latter will conduct the search and return the corresponding encrypted documents to the data user.

The basic idea of our scheme is to modify a key-policy attributed-based encryption (KP-ABE) scheme constructed from bilinear pairing over prime-order groups. Without loss of generality, we will use the large universe KP-ABE scheme selectively secure in the standard model proposed by Rouse- lakis and Waters in [18] to illustrate our construction during the rest of the paper. In KP-ABE, a ciphertext is computed with respect to a set of attributes and an access policy is encoded into a user's private key. A ciphertext can be decrypted by a private key only if the set of attributes associ- ated with the ciphertext satisfies the access policy associated with the private key. Access policies in [18] can be very expressive, supporting any monotonic Boolean formulas. At first sight, a KP-ABE scheme can be transformed to an expressive SE scheme by treating attributes as keywords to be searched, by directly transforming the key generation algorithm on attribute access structures to a trapdoor gener- ation algorithm on keyword search predicates, and by using the decryption algorithm to test whether keywords in a ciphertext satisfy the predicate in a trapdoor. However, KP- ABE schemes (e.g., [18], [19]) are not designed to preserve privacy of attributes (keywords) associated with ciphertexts. Specifically, given the public parameter and a ciphertext, the attributes (keywords) in the ciphertext can be discerned by anyone. In the following, to keep our description compact and consistent, we will use access structure, policy and predicate interchangeably.

In order to hide keywords in a ciphertext, inspired by the "linear splitting" technique in [20], we firstly split ci- phertext components corresponding to every keyword into two randomized complementary components. Thus, even though the ciphertext still contains information about the keywords, this information is computationally infeasible to obtain from the public parameter and the ciphertext. We secondly re-randomize trapdoor components corresponding to every keyword associated with an access structure to match the splitted components in the ciphertext.

In addition to hiding keywords in ciphertexts, we also need to preserve keyword privacy in a trapdoor which contains an access structure as a component. First, to p- reserve keyword privacy in an access structure, we adopt the method in [21] to divide each keyword into a generic name and a keyword value. Since keyword values are much more sensitive than the generic keyword names, the key- word values in an access structure are not disclosed to the cloud server, whereas a partial hidden access structure with only generic keyword names is included in a trapdoor and sent to the cloud server. Take the aforementioned keyword access structure ("Illness = Diabetes" AND ("Age = 30" OR "Weight = 150-200")) as an instance, "Illness", "Age" and "Weight" are the generic names whilst "Diabetes", "30" and "200" are the keyword values. Consequently, the partial hidden access structure ("Illness" AND "Age" OR "Weight") is included in the trapdoor. Second, as in all the PEKS schemes, trapdoors are subject to the offline keyword dictionary guessing attacks. That is, anyone who knows a trapdoor and the public parameter may discover the keyword values embedded in the trapdoor by launching exhaustive searching attacks on keyword values. As a rem- edy to such attacks, we assign a designated cloud server as introduced in [22] to perform the searching operations. We equip this designated server with a public and private key pair of which the public key will be used in trapdoor generation such that it is computationally infeasible for anyone without knowledge of the privacy key to derive keywords information from the trapdoor. Thus, trapdoors can be delivered to the cloud server over a public channel.

We define a security model for expressive SE, which takes into account all adversarial capabilities of the standard SE security notion. The adversary is able to learn trapdoors over access structures of its choice, but it should not be able to learn any information about the keyword values in the challenge ciphertext. Note that since the Rouselakis-Waters KP-ABE scheme [18], which the proposed SE scheme isbuilt upon, is selectively secure, our expressive SE scheme can only be proved to be selectively secure where the adversary has to commit the challenge keyword set in advance.

Contributions

Below we briefly summarize our contributions in this paper.

- We propose the first expressive SE scheme in the public-key setting from bilinear pairings in *prime- order* groups. As such, our scheme is not only capable of expressive multi-keyword search, but also signif- icantly more efficient than existing schemes built in composite-order groups.
- Using a randomness splitting technique, our scheme achieves security against offline keyword dictionary guessing attacks to the ciphertexts. Moreover, to preserve the privacy of keywords against offline keyword dictionary guessing attacks to trapdoors, we divide each keyword into keyword name and

keyword value and assign a designated cloud server to conduct search operations in our construction.

- We formalize the security definition of expressive SE, and formally prove that our proposed expressive SE scheme is selectively secure in the standard model.
- We implement our scheme using a rapidly proto- typing tool called Charm, and conduct extensive experiments to evaluate its performance. Our results confirm that the proposed scheme is sufficiently effi- cient to be applied in practice.

Related Work

Public-Key Encryption with Keyword Search. After Boneh et al. [7] initiated the study of public-key encryption with keyword search (PEKS), several PEKS constructions were put forth using different techniques or considering differ- ent situations [8], [11], [12], [13], [14], [15], [22], [23], [24], [25], [26], [27], [28], [29]. They aim to solve two cruces in PEKS: (1) how to make PEKS secure against offline keyword dictionary guessing attacks; and (2) how to achieve expressive searching predicates in PEKS. In terms of the offline keyword dictionary guessing attacks, which requires that no adversary (including the cloud searching server) can learn keywords from a given trapdoor, to the best of our knowledge, such a security notion is very hard to be achieved in the public-key setting [30]. Regarding the expressive search, there are only few works in PEKS [8], [13], [14], [15]. Unfortunately, the construction in [13] is built on the basis of inner-product predicate encryption [16], and the constructions in [8], [14], [15] are built from the pairings in composite-order group. Therefore, they are not sufficiently efficient to be adopted in the practical world [16], [17].

Moreover, the number of keywords allowed in these search- able schemes are predefined in the system setup phase. We compare our scheme to other keyword search schemes in Table 1. It is straightforward to see that compared to the existing ones, our construction make a good balance in that it allows unbounded keywords, supports expressive access structures, and is built in the prime-order groups.

Private-key Searchable Encryption. In a private-key SE setting, a user uploads its private data to a remote database and keeps the data private from the remote database ad- ministrator. Private-key SE allows the user to retrieve all the records containing a particular keyword from the re- mote database [1], [2], [3]. However, as the name suggests, private-key SE solutions only apply to scenarios where data owners and data users totally trusted each other.

Private Information Retrieval. With respect to public database such as stock quotes, where the user is unaware of it and wishes to search for some data-item without revealing to the database administrator which item it is, private infor- mation retrieval (PIR) [4], [5], [6] protocols were introduced, which allow a user to retrieve data from a public database with far smaller communication then just downloading the entire database. Nevertheless, in our context, the database is not publicly available, the data is not public, so the PIR solutions cannot be applied.

Organization

The remainder of this paper is organized as follows. In Sec- tion 2, we briefly review some of the notions and definitions to be used in the paper. In Section 3, after depicting the sys- tem architecture for our expressive keyword search system, we give a concrete expressive keyword search scheme. In Section 4, we discuss the properties and several extensions of our expressive keyword search scheme. We implement our scheme and compare it with related works in Section 5. We conclude the paper in Section 6.

2 PRELIMINARIES

In this section, we review some basic cryptographic notions and definitions that are to be used later.

Bilinear Pairings and Complexity Assumptions

Let tt be a group of prime order p with a generator g. Let $\hat{e}: tt \times tt \mapsto$ be an efficiently computable bilinear pairing function satisfying the following properties [31].

- Bilinear: for all $g \in tt$, and $a, b \in Z_p^*$, we have $\hat{e}(g^a, g^b) = \hat{e}(g, g)^{ab}$.
- Non-degenerate: $\hat{e}(g, g) \neq 1$.

Decisional Bilinear Diffie-Hellman Assumption [31]. The decisional Bilinear Diffie-Hellman (BDH) problem is that for any probabilistic polynomial-time algorithm, given g, g^a, g^b, g^c, it is difficult to distinguish $(g, g^a, g^b, g^c, \hat{e}(g, g)^{abc})$ from (g, g^a, g^b, g^c, Z), where $g \in tt, Z \in tt$, $a, b, c \in Z_p^*$ chosen independently and uniformly at random.

Decisional $(q-2)$ Assumption [18]. Let q be an integer. The decisional $(q-2)$ problem is that for any probabilistic polynomial-time algorithm, given $\vec{A} = g, g^x, g^y, g^z, g^{(xz)^2}$, $g^{b_i}, g^{xzb_i}, g^{xz/b_i}, g^{x^2zb_i}, g^{y/b^2}, g^{y^2/b^2}$ $\forall i \in [q]$, $g^{xzb_i/b_j}, g^{yb_i/b^2}, g^{xyzb_i/b_j}, g^{(xz)^2b_i/b_j}$ $\forall i, j \in [q], i \neq j$, it is difficult to distinguish $(\vec{A}, \hat{e}(g, g)^{xyz})$ from $(\vec{A}, Z)$, where $g \in tt, Z \in tt, x, y, z, b_1, ..., b_q \in Z_p^*$ chosen independently and uniformly at random.

Decisional Linear Assumption [32]. The decisional lin- ear problem is that for any probabilistic polynomial-time algorithm, given $g, g^{x_1}, g^{x_2}, g^{x_1x_3}, g^{x_2x_4}$, it is difficult to distinguish $(g, g^{x_1}, g^{x_2}, g^{x_1x_3}, g^{x_2x_4}, g^{x_3+x_4})$ from $(g, g^{x_1}, g^{x_2}, g^{x_1x_3}, g^{x_2x_4}, Z)$, where $g, Z \in tt, x_1, x_2, x_3, x_4 \in Z_p^*$ chosen independently and uniformly at random.

Access Structures and Linear Secret Sharing

Following the definition in [33], [34], we describe the notions of access structures and linear secret sharing schemes.

Definition 1. (Access Structure). Let $\{P_1, ..., P_n\}$ be a set of parties. A collection $A \subseteq 2^{\{P_1,...,P_n\}}$ is monotone if $\forall B, C:$ if $B \in A$ and $B \subseteq C$, then $C \in A$. An (monotone) access structure is a (monotone) collection A of non-empty subsets of $\{P_1, ..., P_n\}$, i.e., $A \subseteq 2^{\{P_1,...,P_n\}} \setminus \{\emptyset\}$. The sets in A are called the authorized sets, and the sets not in A are called the unauthorized sets.

In our construction, we only consider monotone access structures. Notice that general access structures in large uni- verse ABE can be realized by splitting the attribute universe in half and treating the attributes of one half as the negated

ISBN 978-93-88808-61-3

TABLE 1
Comparisons of expressive keyword search schemes.

	Keyword Privacy	Expressiveness	Bilinear Group	Security	Unbounded keywords
BCOP04 [7]	keyword guessing attacks on trapdoors	AND	prime	full random oracle	yes
KSW13 [16]	keyword guessing attacks on trapdoors	AND, OR	composite	full standard model	no
LZDLC13 [8]	keyword guessing attacks on trapdoors	AND, OR	composite	full standard model	no
LHZF14 [14]	no keyword guessing attacks on trapdoors	AND, OR, NOT	composite	full standard model	no
Our scheme	keyword guessing attacks on trapdoors by designated server only	AND, OR	prime	selective standard model	yes

(NOT) versions of the attributes in the other half [35]. Also, it has been presented in [36], [37] how to describe non- monotonic access structures in terms of monotonic access structures with negative (NOT) shares.

Definition 2. (Linear Secret Sharing Schemes). Let P be a set of parties. Let M be a matrix of size $l \times$. Let $\rho : 1, \{ ..., l \} \rightarrow P$ be a function that maps a row to a party for labeling. A secret sharing scheme Π over a set of parties P is a linear secret-sharing scheme (LSSS) over Z_p if

1) The shares for each party form a vector over Z_p.
2) There exists a matrix M which has l rows and n columns called the share-generating matrix for Π. For $i = 1, ..., l$, the x-th row of matrix M is labeled

by a party $\rho(i)$, where $\rho : \{1, ..., l\} \rightarrow P$ is a function that maps a row to a party for labeling. Considering that the column vector $\overrightarrow{v} = (\mu, r_2, ..., r_n)$, where $\mu \in Z_p$ is the secret to be shared and $r_2, ..., r_n \in Z_p$ are randomly chosen, then $M\overrightarrow{v}$ is the vector of l shares of the secret μ according to Π. The share $(M\overrightarrow{v})_i$ belongs to party $\rho(i)$.

It has been noted in [33] that every LSSS also enjoys the linear reconstruction property. Suppose that $\Pi$is an LSSS for an access structure A. Let **A** be an authorized set, anddefine $I \subseteq \{1, ..., l\}$ as $I = \{i | \rho(i) \in \mathbf{A}\}$. Then the vector $(1, 0, ..., 0)$ is in the span of rows of M indexed by I, and there exist constants $\{w_i \in Z_p\}_{i \in I}$ such that, for any valid shares $\{v_i\}$ of a secret μ according to Π, we have $\sum_{i \in I} w_i v_i = \mu$. These constants $\{w_i\}$ can be found in the polynomial time in terms of the size of the share-generating matrix M [38]. On the other hand, for an unauthorized set $\mathbf{A}^J$, no such constants $\{w_i\}$ exist. Also, in this case it is true that if $I^J = \{i | \rho(i) \in \mathbf{A}^J\}$, there exists a vector w such that its first component w is any non zero element in Z_p and $< M_i, w > = 0$ for all $i \in I^J$, where M_i is the i-th row of M [18].

Boolean Formulas [33]. Access structures can also be described in terms of monotonic boolean formulas. LSSS access structures are more general, and can be derived from representations as boolean formulas. There are techniques to convert any monotonic boolean formula into a correspond- ing LSSS matrix[3]. The boolean formula can be represented as an access tree, where the interior nodes are AND and OR gates, and the leaf nodes denote attributes. The number of the rows in the corresponding LSSS matrix will be the same as the number of the leaf nodes in the access tree.

3. We give an example on how to convert a boolean formula into an equivalent LSSS matrix in Appendix C.

3 Efficient and Expressive Keyword Search with Unbounded Keywords

In this section, we describe the system model, design goals, threat model and algorithms of our expressive SE scheme.

System Model and Design Goals

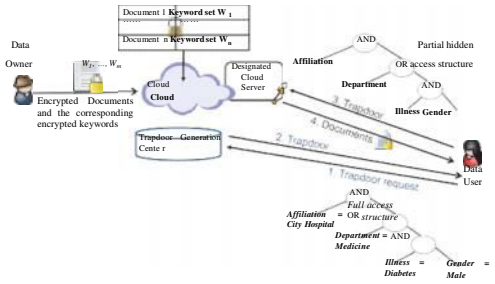


Fig. 1. Architecture of expressive keyword search system.

The architecture of our keyword search system is shown in Fig. 1, which is composed of four entities: a trusted trapdoor generation centre who publishes the system pa- rameter and holds a master private key and is responsible for trapdoor generation for the system, data owners who outsource encrypted data to a public cloud, data users who are privileged to search and access encrypted data, and a designated cloud server who executes the keyword search operations for data users. To enable the cloud server to search over ciphertexts, the data owners append every encrypted document with encrypted keywords[4]. A data user issues a trapdoor request by sending a keyword access structure to the trapdoor generation centre which generates and returns a trapdoor corresponding to the access struc- ture. We assume that the trapdoor generation centre has a separate authentication mechanism to verify each data user and then issue them the corresponding trapdoors. After

4. Note that each keyword is composed of a generic name and a keyword value.

obtaining a trapdoor, the data user sends the trapdoor and the corresponding partial hidden access structure (i.e., the access structure without keyword values) to the designated cloud server. The latter performs the testing operations between each ciphertext and the trapdoor using its private key, and forwards the matching ciphertexts to the data user. As mentioned earlier, a ciphertext created by a data owner consists of two parts: the encrypted document generated using an encryption scheme and the encrypted keywords generated using our SE scheme. From now on, we only consider the latter part of the encrypted document, and ignore the first part since it is out of the scope of this paper.

In summary, the design goals of our expressive SE scheme are fourfold.

- **Expressiveness.** The proposed scheme should sup- port keyword access structures expressed in any Boolean formula with AND and OR gates.
- **Efficiency.** The proposed scheme should be ade- quately efficient in terms of computation, commu- nication and storage for practical applications.
- **Keyword privacy.** First, a ciphertext without its corresponding trapdoors should not disclose any information about the keyword values it contains to the cloud server and outsiders. Second, a trapdoor should not leak information on keyword values to any outside attackers without the private key of the designated cloud server. We capture this notion of security for the SE scheme in terms of semantic security to ensure that encrypted data does not reveal any information about the keyword values, which we call "selective indistinguishability against chosen keyword-set attack (selective IND-CKA security)" (See Appendix A).
- **Provable security.** The security of the proposed scheme should be formally proved under the stan- dard model rather than the informal analysis.

Threat Model

We assume that the trapdoor generation centre is a trusted entity.

The cloud server is assumed to be "honest-but-curious", i.e., it will honestly follow the protocol but it is also curious to learn any private information from the data stored in the cloud. Data owners are assumed to honestly store their data, while data users are not trusted, and they can even collude with a malignant cloud server in order to discover private information of other parties. We assume that the trusted trapdoor generation centre is equipped with a separate authentication mechanism to verify data users before issuing trapdoors to users. Also, we assume that all adversaries have bounded computational capability, so they cannot break the aforementioned difficult problems.

Construction

In the system, the trusted trapdoor generation centre is given a public parameter and a master private key generated by the Setup algorithm, and uses the Trapdoor algorithm to generate a trapdoor T_{M} for some keyword set associated with an access structure $(\mathrm{M}, \rho, \{W_{\rho(i)}\})$ at the request of a privileged data user, where M is an access matrix, ρ is the function that associates the rows of M to the generic names of keywords, and $\{W_{\rho(i)}\}$ are the corresponding keyword values[5]. The cloud server is given a public and private key pair created by the sKeyGen algorithm, and will input the trapdoor given by a data user and its private key to the Test algorithm to determine whether a document contains the keywords satisfying the keyword access structure $(\mathrm{M}, \rho, \{W_{\rho(i)}\})$ specified by the data user.

Let tt be a group of prime order p with a generator g, and $\hat{e}: tt \times tt \rightarrow tt_1$ be the bilinear map. On the basis of the KP-ABE scheme proposed by Rouselakis and Waters [18], which we will refer to as the Rouselakis-Waters KP- ABE scheme, we describe our expressive and unbounded SE system in the prime-order groups as follows.

- Setup. This algorithm takes the security parameter 1^λ as input. It randomly chooses a group tt of prime order p, a generator g and random group elements $u, h, w \in tt$. Also, it randomly chooses $\alpha, d_1, d_2, d_3, d_4 \in Z_p$, and computes $g_1 = g^{d_1}$, $g_2 = g^{d_2}$, $g_3 = g^{d_3}$, $g_4 = g^{d_4}$. Finally, it publishes the public parameter $pars = (H, g, u, h, w, g_1, g_2, g_3, g_4, \hat{e}(g, g)^\alpha)$, where H is a collision-resistant hash function that maps elements in tt_1 to elements in tt, and keeps the master private key $msk = (\alpha, d_1, d_2, d_3, d_4)$.
- sKeyGen. This algorithm takes the public parameter $pars$ as input. It randomly chooses $\gamma \in Z_p^*$, and outputs the public and private key pair $(pk_s, sk_s) = (g^\gamma, \gamma)$ for the server.
- Trapdoor. This algorithm takes the public param- eter $pars$, the server public key pk_s, the master private key msk and an LSSS access structure $(\mathrm{M}, \rho, \{W_{\rho(i)}\})$[6] as input, where M is an $l \times n$ matrix over Z_p, the function ρ associates the rows of M to generic keyword names, and $\{W_{\rho(i)}\}$ are the corre- sponding keyword values. Let M_i be the i-th row of M for $i \in \{1, \ldots, l\}$ and $\rho(i)$ be the keyword name associated with this row by the mapping ρ. It randomly chooses a vector $\vec{y} = (\alpha, y_2, \ldots, y_n)^\perp$ where $y_2, \ldots, y_n \in Z_p$, $r, r^{\mathrm{J}} \in Z_p$, $t_{1,1}, t_{1,2}, \ldots, t_{l,1}, t_{l,2} \in Z_p$, computes $T = g^r$, $T^{\mathrm{J}} = g^{r^{\mathrm{t}}}$, and outputs the trapdoor $T_{\mathrm{M},\rho} = (\mathrm{M}, \rho), T, T^{\mathrm{J}}, \{T_{i,1}, T_{i,2}, T_{i,3}, T_{i,4}, T_{i,5}, T_{i,6}\}_{i \in [1,l]}$ as
$T_{i,1} = g^{v_i} w^{d_1 d_2 t_{i,1} + d_3 d_4 t_{i,2}}$,
$T_{i,2} = H(\hat{e}(pk_s, T^{\mathrm{J}})^r) \cdot g^{d_1 d_2 t_{i,1} + d_3 d_4 t_{i,2}}$,
$T_{i,3} = ((u^{W_{\rho(i)}} h)^{t_{i,1}})^{-d_2}$, $T_{i,4} = ((u^{W_{\rho(i)}} h)^{t_{i,1}})^{-d_1}$,
$T_{i,5} = ((u^{W_{\rho(i)}} h)^{t_{i,2}})^{-d_4}$, $T_{i,6} = ((u^{W_{\rho(i)}} h)^{t_{i,2}})^{-d_3}$,
where $v_i = \mathrm{M}_i \cdot \vec{y}$ is the share associated with the row M_i of the access matrix M. Note that only (M, ρ) is included in the trapdoor $T_{\mathrm{M},\rho}$.
- Encrypt. This algorithm takes the public parameter $pars$ and a keyword set $\mathbf{W}$ (each keyword is denoted as $N_i = W_i$, where N_i is the generic keyword name and W_i is the corresponding keyword value) as in- put. Let m be the size of $\mathbf{W}$, and $W_1, \ldots, W_m \in Z_p$ be

5. To be distinguishable from the keyword value W_i in a ciphertxt, we use $W_{\rho(i)}$ to denote the keyword value in a trapdoor.

6. For the details about how to convert a boolean formula into an equivalent LSSS matrix, please refer to [33].

ISBN 978-93-88808-61-3

the values of **W**. It randomly chooses $\mu, s_{1,1}, s_{1,2}, \ldots, s_{m,1}, s_{m,2}, z_1, \ldots, z_m \in Z_p$, and outputs a ciphertext CT = $C, D, \{(D_i, E_{i,1}, E_{i,2}, F_{i,1}, F_{i,2})\}_{i\in[1,m]}$ as

$$C = \hat{e}(g, g)^{\alpha\mu}, \quad D = g^{\mu},$$
$$D_i = w^{-\mu}(u^{W_i}h)^{z_i}, \quad E_{i,1} = g_1^{z_i - s_{i,1}},$$
$$E_{i,2} = g_2^{s_{i,1}}, \quad F_{i,1} = g_3^{z_i - s_{i,2}}, \quad F_{i,2} = g_4^{s_{i,2}}.$$

Note that in the implementation, to efficiently conduct keyword search, the ciphertext will be stored along with the generic names $\{N_i\}$ corresponding to keyword values $\{W_i\}$. Thus, before performing the Test algorithm on the encrypted keyword values, the matching on the keyword names will be executed, thereby reducing the searching time.

- Test. This algorithm takes the public parameter *pars*, the server private key sk_S, a ciphertext $C, D, \{(D_i, E_{i,1}, E_{i,2}, F_{i,1}, F_{i,2})\}$ on a keyword set **W** and a trapdoor $T_{M,\rho}$ associated with an access structure (M, ρ, $\{W_{\rho(i)}\}$) as input. It calculates $I_{M,\rho}$ from (M, ρ), which is a set of minimum subsets satisfying (M, ρ). It then checks whether there is an $\mathbf{I} \in I_{M,\rho}$ satisfying

$$\prod_{i\in\mathbf{I}} \left(\hat{e}(D, T_{i,1})\hat{e}\left(D_i, \frac{T_{i,2}}{H(\hat{e}(T, T^{J})^{\gamma})}\right)\hat{e}(E_{i,1}, T_{i,3})\right.$$
$$\left.\hat{e}(E_{i,2}, T_{i,4})\hat{e}(F_{i,1}, T_{i,5})\hat{e}(F_{i,2}, T_{i,6})\right)^{\Sigma_{w_i}} = C,$$

where $\sum_{i\in\mathbf{I}} w_i \mathbf{M}_i = (1, 0, \ldots, 0)$. It outputs 0 if no element in $I_{M,\rho}$ satisfying this equation or 1 otherwise.

Remarks. In our construction, the term g^{z_i} in the original construction [18] is split into $g^{z_i - s_{i,1}}$ and $g^{s_{i,1}}$. Thus, W_i is subtly hidden from the ciphertext. To see this, we have

$$\hat{e}(D_i, g_1) = \hat{e}(w, D)^{-1} \cdot \hat{e}(u^{W_i}h, g^{r_i}),$$

where the term $g_1^{z_i}$ is embedded in $g_1^{z_i - s_{i,1}}$, which cannot be computed from $g_1^{z_i - s_{i,1}}$ without knowing the value of $g_1^{s_{i,1}}$. Nevertheless, given $g_2^{s_{i,1}}$, it is difficult to compute the value of $g_1^{s_{i,1}}$. Similarly, the result works with $g_3^{z_i - s_{i,2}}$, $g_4^{s_{i,2}}$ as well. Note that some redundant elements as g_3, g_4, $T_{i,5}$, $T_{i,6}$, $F_{i,1}$, $F_{i,2}$ are introduced in our scheme to make the proof go smoothly.

Correctness

If the keyword set **W** embedded in a ciphertext satisfies the access structure associated with the trapdoor, we will have $\sum_{i\in\mathbf{I}} v_i w_i = \alpha$. Therefore,

$$\prod_{i\in\mathbf{I}} \left(\hat{e}(D, T_{i,1})\hat{e}\left(D_i, \frac{T_{i,2}}{H(\hat{e}(T, T^{J})^{\gamma})}\right)\hat{e}(E_{i,1}, T_{i,3})\right.$$
$$\left.\hat{e}(E_{i,2}, T_{i,4})\hat{e}(F_{i,1}, T_{i,5})\hat{e}(F_{i,2}, T_{i,6})\right)^{\Sigma_{w_i}}$$
$$= \prod_{i\in\mathbf{I}} \hat{e}(g^{\mu}, g^{v_i w^{d_1 d_2 t_{i,1} + d_3 d_4 t_{i,2}}})^{w_i}$$
$$\cdot \hat{e}(w^{-\mu}(u^{W_i}h)^{z_i}, g^{d_1 d_2 t_{i,1} + d_3 d_4 t_{i,2}})^{w_i}$$
$$\cdot \hat{e}(g_1^{z_i - s_{i,1}}, ((u^{W_{\rho(i)}}h)^{t_{i,1}})^{-d^2})^{w_i}$$
$$\cdot \hat{e}(g_2^{s_{i,1}}, ((u^{W_{\rho(i)}}h)^{t_{i,1}})^{-d^1})^{w_i}$$
$$\cdot \hat{e}(g_3^{z_i - s_{i,2}}, ((u^{W_{\rho(i)}}h)^{t_{i,2}})^{-d^4})^{w_i}$$
$$\cdot \hat{e}(g_4^{s_{i,2}}, ((u^{W_{\rho(i)}}h)^{t_{i,2}})^{-d^3})^{w_i}$$
$$= \hat{e}(g, g)^{\mu \sum_{i\in\mathbf{I}} v_i w_i} = \hat{e}(g, g)^{\alpha\mu}$$

Security Proof

Theorem 1. Under the decisional BDH assumption, the $(q - 2)$ assumption and the decisional linear assumption, our scheme is selectively indistinguishable under chosen keyword-set attacks (selective IND-CKA security).

Proof. The details of the selective IND-CKA security definition and its proof are given in Appendix B. The proof is divided into two parts, depending on the role of the adversary. In the first part, the adversary is assumed to be an outside attacker, and in the second part, the adversary is assumed to be the cloud server who performs search operations.

4 Discussion and Analysis

In this section, we discuss the properties as well as extensions of our expressive SE scheme.

Keyword Privacy

Keyword Value Guessing Attacks on Ciphertexts. Below we briefly review the encryption algorithm of the KP-ABE scheme in [18], and then show that there exists a keyword value guessing attack if it is directly transformed into a searchable encryption scheme.

Encrypt. Let n denote the size of **W** and $W_1, \ldots, W_n$ be the specific values of **W**. It randomly chooses $\mu, z_1, \ldots, z_m \in Z_p$, and outputs a ciphertext CT = $C, D, \{(C_i, D_i)\}_{i\in[1,m]}$.

$$C = \hat{e}(g, g)^{\alpha\mu}, \quad D = g^{\mu},$$
$$\forall i \in [m] \quad C_i = w^{-\mu}(u^{W_i}h)^{z_i}, \quad D_i = g^{z_i},$$

where $g, u, h, w, \hat{e}(g, g)^{\alpha}$ are the public parameters.

Given a ciphertext CT = $C, D, \{(C_i, D_i)\}_{i\in[1,m]}$, an adversary can easily determine whether a keyword value W^{J} is incorporated in the ciphertext by checking whether the following equation holds.

$$\hat{e}(C_i, g) = \hat{e}(w^{-1}, D) \cdot \hat{e}(u^{W^{t}}h, D_i).$$

In order to prevent such attacks, in our construction, we use a "linear splitting" technique [20] on each keyword value related component of the ciphertext, and then re-randomize the components upon each keyword value in the trapdoor. The former step prevents keyword value guessing attacks to the ciphertext while the latter step allows the trapdoor to be used for testing keyword values in the ciphertext.

Keyword Value Guessing Attacks on Trapdoors. Concerning this security requirement, we need to tackle two problems in our construction. First, keywords associated with a trapdoor must be hidden from the access structure. We address this problem by separating each keyword into a generic name and a keyword value, i.e., each keyword is in the form of "generic name = keyword value", and a partial hidden access structure, i.e., the full access structure with keyword values being removed (See Fig. 1) is incorporated in a trapdoor and given to the designated cloud server. Second, the entire trapdoor should be immune to the offline keyword value guessing attacks [25]. In our SE system,

we resort to a weaker security notion by requiring that a trapdoor will not disclose information about the keyword values in the ciphertext to an adversary excluding the cloud server who executes the searching operations. We assign a designated cloud server [22] to conduct search and equip it with a public and private key pair. Since the components in a trapdoor are tied with the public key of the server, only the designated cloud server with the corresponding private key is capable to learn the keyword values hidden in the trapdoor by performing offline guessing attacks.

Unbounded keyword search

In "small universe" KP-ABE constructions [18], the size of the keyword space were polynomially bounded in the security parameter and the keywords were fixed at the setup phase. Moreover, the sizes of the public parameter- s grow linearly with the number of keywords [8], [14], [15]. On the contrary, in "large universe" constructions, the size of the keyword space can be exponentially large, so it is much more desirable in the real-world applications. Our construction of the expressive SE scheme inherits the advantages of the Rouselakis-Waters scheme [18]. Thus, it is straightforward to see that in our SE scheme, the size of the public parameter is immutable with the number of keywords, and the number of the keywords allowed for the system is unlimited and can be freely set.

Extensions

Our expressive SE system can be extended in several ways.

- Expressive searchable encryption for the range search. Range search is an important requiremen- t for searchable encryption in many applications. By defining keywords in a hierarchical manner as shown in [27], we can directly expand our SE system to support a class of simple range search [27]. Take a keyword name "Age" with keyword values from 0 to 100 as an example. The path of the leaf node "11-20" is ("0-100", "0-30", "11-20"), and "0-30", "0-10" are simple ranges from level-2 and level-3, respectively.
- Anonymous KP-ABE. Our SE system is built by anonymizing the Rouselakis-Waters KP-ABE scheme [18]. Therefore, our scheme can be easily extended to obtain an unbounded and anonymous KP-ABE scheme in the prime-order group without random oracles, in which an adversary, given a ciphertext, cannot learn any information about the associated attribute set.
- Anonymous hierarchical identity-based encryption (HIBE). The Rouselakis-Waters KP-ABE scheme in [18] can be converted to an HIBE scheme using non-repeating identities, "AND" policies and delegation capabilities [19]. Since our SE scheme can be used to construct an anonymous KP-ABE scheme, it can be further converted to an anonymous HIBE scheme using the same method as in [19].

5 Performance Analysis

We implement our construction of expressive SE in the prime-order group in Charm [39], which is a programming environment for rapid prototyping of cryptographic prim- itives. In this section, we compare the computational cost, communication and storage overhead of our scheme with other existing schemes.

Comparison

Let $|pars|$, $|msk|$, $|CT|$, $|T_{M,\rho}|$, $|M|$ be the sizes of the public parameter, the master private key, the ciphertext, the trapdoor and the access structure, respectively. Let k be the length of the vector corresponding to the ciphertext in [16], l be the number of keywords in an access structure, n be the maximum number of keywords allowed for the system, and m be the size of a keyword set ascribed to a ciphertext. Denote E as an exponentiation operation, P as a pairing operation, χ_1 as the number of elements in $I_{M,\rho} = \{I_1, ..., I_{\chi_1}\}$, χ_2 as $|I_1| + ... + |I_{\chi_1}|$, and χ_3 as the number of primed keywords [14] in a search predicate.

TABLE 2
Comparison of Storage and Communication Overhead

	Public parameter $\|pars\|$	Master private key $\|msk\|$	Trap-door $\|T_{M,\rho}\|$	Cipher-text $\|CT\|$
KSW13 [16]	$2k+3$	$2k+4$	$2k+1+\|M\|$	$2k+1$
LZDLC13 [8]	$n+5$	$n+4$	$2l+\|M\|$	$m+2$
LHZF14 [14]	$n+4$	$n+2$	$3l+\|M\|$	$m+2$
Our Scheme	9	5	$6l+\|M\|$	$5m+2$

We compare our searchable encryption system with the other three known expressive SE schemes [8], [14], [16] in Table 2 which are all constructed over composite order groups. From Table 2, it is not difficult to see that our construction is the only one that supports unbounded num- ber of keywords in the expressive keyword search systems. Note that our scheme is measured in terms of number of elements in prime order groups while the other three schemes are measured in terms of number of elements in composite order groups. According to the analysis in [40][7], in terms of the pairing-friendly elliptic curves, prime order groups have a clear advantage in the parameter sizes over composite order groups.

TABLE 3
Comparison of Computation Overhead.

	KSW13 [16]	LZDLC13 [8]	LHZF14 [14]	Ours
Trap-door	$6k \cdot E$	$4l \cdot E$	$4l \cdot E$	$16l \cdot E + E$
Enc.	$4k \cdot E + E$	$2(m+1) \cdot E$	$(m+2) \cdot E + P$	$7m \cdot E + 2 \cdot E$
Test	$2k \cdot P + P$	$\leq \chi_2 \cdot E + 2\chi_2 \cdot P$	$\leq \chi_2 \cdot E + 2\chi_2 \cdot P + 2\chi_3 \cdot P$	$\leq \chi_2 \cdot E + E + P + 6\chi_2 \cdot P$
Group Order	Composite	Composite	Composite	Prime

In Table 3, we compare the computational costs incurred in the systems from [8], [14], [16] and our system. It is worth noticing that as mentioned in [17], "a Tate pairing on a 1024- bit composite-order elliptic curve is roughly 50 times slower than the same pairing on a comparable prime-order curve,

7. See Table 3 in [40] for the results.

and this performance gap will only get worse at higher security levels". Therefore, although our SE system requires more exponentiation and pairing operations than the other systems, it is far more computationally efficient than the other three schemes.

Experimental Results

We implement our scheme in Charm [39][8], which is a framework developed to facilitate rapid prototyping of cryptographic schemes and protocols. Based on the Python programming language, Charm enables one to implement a cryptographic scheme with very few lines of code, signif- icantly reducing development time. Meanwhile, computa- tionally intensive mathematical operations are implement- ed with native modules, so the overhead due to Python in Charm is less than 1%. Since all Charm routines are designed under the asymmetric groups, our construction is transformed to the asymmetric setting before the imple- mentation. That is, three groups tt, $\hat{tt}$ and tt_1 are used and the pairing $\hat{e}$ is a function from $tt \times \hat{tt}$ to tt_1. Notice that it has been stated in [18] that the assumptions and the security proofs can be converted to the asymmetric setting in a generic way.

We use Charm of version charm-0.43 and Python 3.4 in our implementation. Along with charm-0.43, we install the latest PBC library for underlying cryptographic operations. Our experiments run on an all-in-one desktop computer with Intel Core i7-4785T CPU (4 core 2.20GHz) and 8GB RAM running 64-bit Ubuntu 15.10.

The computational costs of the Setup and sKeyGen algo- rithms are straightforward, and we focus on the computa- tional costs of the Trapdoor, Encrypt and Test algorithms. In our experiments, a set of keywords is generated, of which every keyword contains a generic name such as "Illness", "Position", "Affiliation" and a keyword value such as "Di- abetes", "Doctor", and "City Hospital". For the sake of simple implementation, we use integers to denote keyword values, e.g., a keyword as "Illness = 6" is expressed by "Illness = Diabetes". In this way, we generate a random set of keywords containing 10 to 50 keywords, and use them to encrypt 5,000 documents. We then remove the keyword values in the ciphertexts such that they contain only generic names of keywords like "Illness", "Position", as specified in our concrete construction.

Thereafter, we randomly choose 2 to 10 keywords to form a random access structure. The number of keywords in a searching query is normally less than 10, according to the searching query logs of search engines [41]. The policy tree is formed such that for any interior node the difference on the node number of its left branch and that of its right branch is less than 2. We generate 50 different access policy trees, 10 for each different number of keywords, and create a trapdoor for each policy tree. We also remove the keyword value information from the trapdoors. So the policy tree in a trapdoor contains only keyword names, e.g., (("Illness" AND "Position") OR "Affiliation").

Also, we take each trapdoor to conduct search over the ciphertexts. For any combination of the keyword names in the ciphertext that satisfies the access policy of the trapdoor, our keyword search scheme runs the Test algorithm to further confirm whether it is an exact match.

All these experiments are conducted over 4 different elliptic curves: SS512, MNT159, MNT201 and MNT224, of which SS512 is a supersingular elliptic curve with the bi- linear pairing on it being symmetric Type 1 pairing, and the pairings on the other three curves are asymmetric Type 3 pairings. These four curves provides security levels of 80- bit, 80-bit, 100-bit and 112-bit, respectively. The computation time for the exponentiation and pairing calculation over the four curves are listed in Fig. 2.

ECs(time in ms)	Exp. tt	Exp. $\hat{tt}$	Exp. tt_1	Pairing
SS512	0.194	0.194	0.027	0.881
MNT159	0.068	0.584	0.160	3.148
MNT201	0.101	0.762	0.207	4.194
MNT224	0.131	0.968	0.252	5.169

Fig. 2. Computational costs for the group operations and pairings over different elliptic curves on a desktop with 2.2GHz 4 core CPU.

Fig. 3 shows the computational overhead for generating trapdoors containing 2 keywords to 10 keywords, from which we can see that the computation time for the trapdoor generation is almost linear to the number of keywords as- sociated with the access structure in the trapdoor. The MNT curves with higher security levels have longer computation time, so MNT224 has higher computation cost among all curves. The computation time of SS512 is close to that of MNT224 due to its higher exponentiation cost over tt. The computation time of generating a trapdoor with 10 keywords is only 0.22s for MNT224, which is quite modest for a powerful trapdoor generation centre.

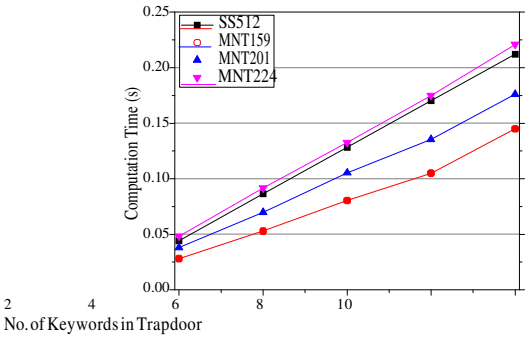


Fig. 3. Computational cost of the Trapdoor algorithm for different curves with respect to number of keywords in trapdoor.

Fig. 4 demonstrates the computation time for the Encrypt algorithm over 10 keywords to 50 keywords. As expected in our analysis, it shows that the computation time is approximately linear to the number of keywords used to generate the ciphertext. The MNT curves with higher secu- rity levels are more expensive in computation cost, while

8. For the explicit information on Charm, please refer to [39]. Note that since it has been clearly shown in [18], [40] that the efficiency of schemes in composite-order groups is much worse than that of schemes in prime-order groups, we will not implement those schemes in composite-order groups (e.g., [8], [14], [16]). In addition, the current version Charm does not support cryptographic schemes in composite- order groups.

the encryption cost of SS512 is much less than that of MNT curves. This is due to the fact that $(4m + 1)$ exponentiations are done in $tt\hat{}$ for the total $(7m + 2)$ exponentiations (see Table 3). To encrypt a document with 50 keywords using MNT224 curve, the computation time is about 1.6s, which is acceptable for most applications.

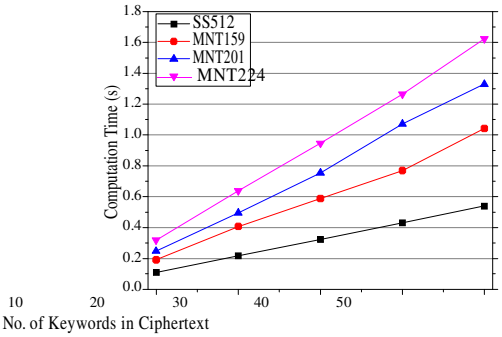


Fig. 4. Computational cost of the Encrypt algorithm for different curves with respect to number of keywords in ciphertext.

The computational cost for the Test algorithm is much more involved. It depends on χ_2, the total number of keywords in all combinations of keywords satisfying the access policy that need to be tried by the cloud server. χ_2 is determined by the access policy and the keywords used to encrypt a document. Fig. 5 shows the relation between the computation time of the Test algorithm and the number of keywords in the access structure of the Trapdoor algorithm. From Fig. 5, it is easy to see that the computation time raises as the number of keywords in the trapdoor and the cipher- text increases. When the trapdoor contains only 2 keywords, the computation time increases quite slowly as keywords in ciphertexts increases. Whilst when the trapdoor has 10 keywords, the computation time grows exponentially as the number of keywords in ciphertexts grows. Among all the curves, SS512 has the best performance, while MNT224 has the highest computational cost. For the 4 curves tested in our experiments, the computation time of searching a document ranges from 50s to 250s for a trapdoor with 10 keywords and a ciphertext with 50 keywords. The computation time can be significantly reduced if keyword search is performed by a powerful cloud server.

6 Conclusions

In order to allow a cloud server to search on encrypted data without learning the underlying plaintexts in the public- key setting, Boneh [7] proposed a cryptographic primitive called public-key encryption with keyword search (PEKS). Since then, considering different requirements in practice, e.g., communication overhead, searching criteria and secu- rity enhancement, various kinds of searchable encryption systems have been put forth. However, there exist only a few public-key searchable encryption systems that support expressive keyword search policies, and they are all built from the inefficient composite-order groups [17]. In this paper, we focused on the design and analysis of public-key searchable encryption systems in the prime-order groups that can be used to search multiple keywords in expressive searching formulas. Based on a large universe key-policy attribute-based encryption scheme given in [18], wepresent- ed an expressive searchable encryption system in the prime- order group which supports expressive access structures expressed in any monotonic Boolean formulas. Also, we proved its security in the standard model, and analyzed its efficiency using computer simulations.

Acknowledgments

This research work is supported by the Singapore Na- tional Research Foundation under the NCR Award Num- ber NRF2014NCR-NCR001-012, and the National Natural Science Foundation of China under the Grant Number 61370027.

References

[1] O. Goldreich and R. Ostrovsky, "Software protection and simula- tion on oblivious rams," *J. ACM*, vol. 43, no. 3, pp. 431–473, 1996.

[2] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in *2000 IEEE Symposium on Security and Privacy, Berkeley, California, USA, May 14-17, 2000*. IEEE Computer Society, 2000, pp. 44–55.

[3] E. Goh, "Secure indexes," *IACR Cryptology ePrint Archive*, vol. 2003, p. 216, 2003.

[4] C. Cachin, S. Micali, and M. Stadler, "Computationally private information retrieval with polylogarithmic communication," in *Advances in Cryptology - EUROCRYPT '99, International Conference on the Theory and Application of Cryptographic Techniques, Prague, Czech Republic, May 2-6, 1999, Proceeding*, ser. Lecture Notes in Computer Science, vol. 1592. Springer, 1999, pp. 402–414.

[5] G. D. Crescenzo, T. Malkin, and R. Ostrovsky, "Single database pri- vate information retrieval implies oblivious transfer," in *Advances in Cryptology - EUROCRYPT 2000, International Conference on the Theory and Application of Cryptographic Techniques, Bruges, Belgium, May 14-18, 2000, Proceeding*, ser. Lecture Notes in Computer Sci- ence, vol. 1807. Springer, 2000, pp. 122–138.

[6] W. Ogata and K. Kurosawa, "Oblivious keyword search," *J. Complexity*, vol. 20, no. 2-3, pp. 356–371, 2004.

[7] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in *Advances in Cryptology - EUROCRYPT 2004, International Conference on the Theory and Applications of Cryptographic Techniques, Interlaken, Switzerland, May 2-6, 2004, Proceedings*, ser. Lecture Notes in Computer Science, vol. 3027. Springer, 2004, pp. 506–522.

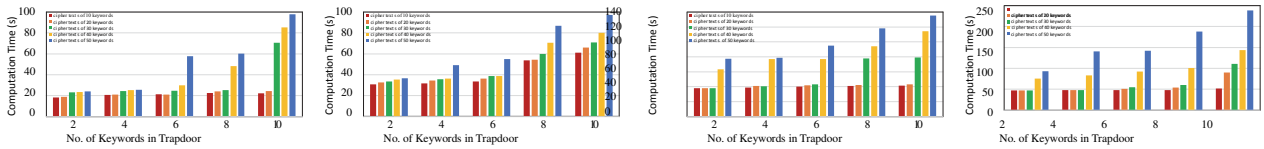
[8] J. Lai, X. Zhou, R. H. Deng, Y. Li, and K. Chen, "Expressive search on encrypted data," in *8th ACM Symposium on Information, Computer and Communications Security, ASIA CCS '13, Hangzhou, China - May 08 - 10, 2013*. ACM, 2013, pp. 243–252.

[9] P. Golle, J. Staddon, and B. R. Waters, "Secure conjunctive key- word search over encrypted data," in *Applied Cryptography and Network Security, Second International Conference, ACNS 2004, Yellow Mountain, China, June 8-11, 2004, Proceedings*, ser. Lecture Notes in Computer Science, vol. 3089. Springer, 2004, pp. 31–45.

[10] D. J. Park, K. Kim, and P. J. Lee, "Public key encryption with conjunctive field keyword search," in *Information Security Applications, 5th International Workshop, WISA 2004, Jeju Island, Korea, August 23- 25, 2004, Revised Selected Papers*, ser. Lecture Notes in Computer Science, vol. 3325. Springer, 2004, pp. 73–86.

[11] Y. H. Hwang and P. J. Lee, "Public key encryption with conjunctive keyword search and its extension to a multi-user system," in *Pairing-Based Cryptography - Pairing 2007, First International Con- ference, Tokyo, Japan, July 2-4, 2007, Proceedings*, ser. Lecture Notes in Computer Science, vol. 4575. Springer, 2007, pp. 2–22.

[12] B. Zhang and F. Zhang, "An efficient public key encryption with conjunctive-subset keywords search," *J. Network and Computer Applications*, vol. 34, no. 1, pp. 262–267, 2011.



(a) SS512 (b) MNT159 (c) MNT201 (d) MNT224

Fig. 5. Experimental results for the Test algorithm over different ellipticcurves.

[13] D. Boneh and B. Waters, "Conjunctive, subset, and range queries on encrypted data," in *Theory of Cryptography, 4th Theory of Cryptography Conference, TCC 2007, Amsterdam, The Netherlands, February 21-24, 2007, Proceedings*, ser. Lecture Notes in Computer Science, vol. 4392. Springer, 2007, pp. 535–554.

[14] Z. Lv, C. Hong, M. Zhang, and D. Feng, "Expressive and secure searchable encryption in the public key setting," in *Information Security - 17th International Conference, ISC 2014, Hong Kong, China, October 12-14, 2014. Proceedings*, ser. Lecture Notes in Computer Science, vol. 8783. Springer, 2014, pp. 364–376.

[15] J. Shi, J. Lai, Y. Li, R. H. Deng, and J. Weng, "Authorized keyword search on encrypted data," in *Computer Security - ESORICS 2014 - 19th European Symposium on Research in Computer Security, Wroclaw, Poland, September 7-11, 2014. Proceedings, Part I*, ser. Lecture Notes in Computer Science, vol. 8712. Springer, 2014, pp. 419–435.

[16] J. Katz, A. Sahai, and B. Waters, "Predicate encryption supporting disjunctions, polynomial equations, and inner products," *J. Cryp- tology*, vol. 26, no. 2, pp. 191–224, 2013.

[17] D. M. Freeman, "Converting pairing-based cryptosystems from composite-order groups to prime-order groups," in *Advances in Cryptology - EUROCRYPT 2010, 29th Annual International Conference on the Theory and Applications of Cryptographic Techniques, French Riviera, May 30 - June 3, 2010. Proceedings*, ser. Lecture Notes in Computer Science, vol. 6110. Springer, 2010, pp. 44–61.

[18] Y. Rouselakis and B. Waters, "Practical constructions and new proof methods for large universe attribute-based encryption," in *2013 ACM SIGSAC Conference on Computer and Communications Security, CCS'13, Berlin, Germany, November 4-8, 2013*. ACM, 2013, pp. 463–474.

[19] A. B. Lewko and B. Waters, "Unbounded HIBE and attribute- based encryption," in *Advances in Cryptology - EUROCRYPT 2011 - 30th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tallinn, Estonia, May 15-19, 2011. Proceedings*, ser. Lecture Notes in Computer Science, vol. 6632, 2011, pp. 547–567.

[20] X. Boyen and B. Waters, "Anonymous hierarchical identity-based encryption (without random oracles)," in *Advances in Cryptology - CRYPTO 2006, 26th Annual International Cryptology Conference, Santa Barbara, California, USA, August 20-24, 2006, Proceedings*, ser. Lecture Notes in Computer Science, vol. 4117. Springer, 2006, pp. 290–307.

[21] J. Lai, R. H. Deng, and Y. Li, "Expressive CP-ABE with partially hidden access structures," in *7th ACM Symposium on Information, Compuer and Communications Security, ASIACCS '12, Seoul, Korea, May 2-4, 2012*. ACM, 2012, pp. 18–19.

[22] H. S. Rhee, J. H. Park, W. Susilo, and D. H. Lee, "Improved searchable public key encryption with designated tester," in *Pro- ceedings of the 2009 ACM Symposium on Information, Computer and Communications Security, ASIACCS 2009, Sydney, Australia, March 10-12, 2009*. ACM, 2009, pp. 376–379.

[23] M. Bellare, A. Boldyreva, and A. O'Neill, "Deterministic and efficiently searchable encryption," in *Advances in Cryptology - CRYPTO 2007, 27th Annual International Cryptology Conference, San- ta Barbara, CA, USA, August 19-23, 2007, Proceedings*, ser. Lecture Notes in Computer Science, vol. 4622. Springer, 2007, pp.535–552.

[24] C. Gu, Y. Zhu, and H. Pan, "Efficient public key encryption with keyword search schemes from pairings," in *Information Security and Cryptology, Third SKLOIS Conference, Inscrypt 2007, Xining, China, August 31 - September 5, 2007, Revised Selected Papers*, ser. Lecture Notes in Computer Science, vol. 4990. Springer, 2007, pp. 372–383.

[25] J. Baek, R. Safavi-Naini, and W. Susilo, "Public key encryption with keyword search revisited," in *Computational Science and Its Applications - ICCSA 2008, International Conference, Perugia, Italy, June 30 - July 3, 2008, Proceedings, Part I*, ser. Lecture Notes in Computer Science, vol. 5072. Springer, 2008, pp. 1249–1259.

[26] Q. Tang and L. Chen, "Public-key encryption with registered key- word search," in *Public Key Infrastructures, Services and Applications - 6th European Workshop, EuroPKI 2009, Pisa, Italy, September 10-11, 2009, Revised Selected Papers*, ser. Lecture Notes in Computer Science, vol. 6391. Springer, 2009, pp. 163–178.

[27] M. Li, S. Yu, N. Cao, and W. Lou, "Authorized private keyword search over encrypted data in cloud computing," in *2011 Inter- national Conference on Distributed Computing Systems, ICDCS 2011, Minneapolis, Minnesota, USA, June 20-24, 2011*. IEEE Computer Society, 2011, pp. 383–392.

[28] H. S. Rhee, J. H. Park, and D. H. Lee, "Generic construction of designated tester public-key encryption with keyword search," *Inf. Sci.*, vol. 205, pp. 93–109, 2012.

[29] W. Yau, R. C. Phan, S. Heng, and B. Goi, "Keyword guessing attacks on secure searchable public key encryption schemes with a designated tester," *Int. J. Comput. Math.*, vol. 90, no. 12, pp.2581– 2587, 2013.

[30] E. Shen, E. Shi, and B. Waters, "Predicate privacy in encryption systems," in *Theory of Cryptography, 6th Theory of Cryptography Conference, TCC 2009, San Francisco, CA, USA, March 15-17, 2009. Proceedings*, ser. Lecture Notes in Computer Science, vol. 5444. Springer, 2009, pp. 457–473.

[31] D. Boneh and M. Franklin, "Identity-based encryption from the weil pairing," in *CRYPTO*, ser. Lecture Notes in Computer Science, vol. 2139. Springer-Verlag, 2001, pp. 213–219.

[32] D. Boneh, X. Boyen, and H. Shacham, "Short group signatures," in *Advances in Cryptology - CRYPTO 2004, 24th Annual International CryptologyConference, Santa Barbara, California, USA, August 15-19, 2004, Proceedings*, ser. Lecture Notes in Computer Science, vol. 3152. Springer, 2004, pp. 41–55.

[33] A. B. Lewko and B. Waters, "Decentralizing attribute-based encryption," in *Advances in Cryptology - EUROCRYPT 2011 - 30th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tallinn, Estonia, May 15-19, 2011. Proceedings*, ser. Lecture Notes in Computer Science, vol. 6632. Springer, 2011, pp. 568–588.

[34] B. Waters, "Ciphertext-policy attribute-based encryption: An expressive, efficient, and provably secure realization," in *Public Key Cryptography - PKC 2011 - 14th International Conference on Practice and Theory in Public Key Cryptography, Taormina, Italy, March 6-9, 2011. Proceedings*, ser. Lecture Notes in Computer Science, vol. 6571. Springer, 2011, pp. 53–70.

[35] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data," in *Proceedings of the 13th ACM Conference on Computer and Communications Security, CCS 2006, Alexandria, VA, USA, Ioctober 30 - November 3, 2006*. ACM, 2006, pp. 89–98.

[36] R. Ostrovsky, A. Sahai, and B. Waters, "Attribute-based encryption with non-monotonic access structures," in *Proceedings of the 2007 ACM Conference on Computer and Communications Security, CCS 2007, Alexandria, Virginia, USA, October 28-31, 2007*. ACM, 2007, pp. 195–203.

[37] A. B. Lewko, A. Sahai, and B. Waters, "Revocation systems with very small private keys," in *31st IEEE Symposium on Security and Privacy, S&P 2010, 16-19 May 2010, Berleley/Oakland, California, USA*. IEEE Computer Society, 2010, pp. 273–285.

[38] A. Beimel, "Secure schemes for secret sharing and key distri- bution," Ph.D. dissertation, Israel Institute of Technology, Israel Institute of Technology, June 1996.

[39] J. A. Akinyele, C. Garman, I. Miers, M. W. Pagano, M. Rushanan, M. Green, and A. D. Rubin, "Charm: a framework for rapidly prototyping cryptosystems," *J. Cryptographic Engineering*, vol. 3, no. 2, pp. 111–128, 2013.

[40] A. Guillevic, "Comparing the pairing efficiency over composite- order and prime-order elliptic curves," in *Applied Cryptography and Network Security - 11th International Conference, ACNS 2013, Banff, AB, Canada, June 25-28, 2013. Proceedings*, ser. Lecture Notes in Computer Science, vol. 7954. Springer, 2013, pp. 357–372.

[41] L. Yang, Q. Mei, K. Zheng, and D. A. Hanauer, "Query log analysis of an electronic health record search engine," in *Proc. of AMIA Annual Symposium*, 2011, p. 915C924.

APPENDIX A
SYSTEM FRAMEWORK AND SECURITY DEFINITION

Our expressive keyword search system consists of setup algorithm Setup, server key generation algorithm sKeyGen, trapdoor generation algorithm Trapdoor, encryption algo- rithm Encrypt and testing algorithm Test.

- Setup(1^λ) $\rightarrow$ (*pars*, *msk*). Taking the security parameter λ as the input, this setup algorithm outputs the public parameter *pars* and the master private key *msk* for the system. This algorithm is run by the trapdoor centre.
- sKeyGen(*pars*) $\rightarrow$ (sk_s, pk_s). Taking the public parameter *pars* as the input, the server key generation algorithm outputs a public and private key pair for the designated searching server. This algorithm is run by the trapdoor centre.
- Trapdoor(*pars*, pk_s, *msk*, (M, ρ,{ $W_{\rho(i)}$ })) $\rightarrow$ T_M. Taking the public parameter *pars*, the server public key pk_s and an access structure (M, ρ{, $W_{\rho(i)}$}) over the universe of keywords as the input, this trapdoor generation algorithm generates a trapdoor T_M. This algorithm is run by the trapdoor centre.
- Encrypt(*pars*, **W**)$\rightarrow$ CT. Taking the public parameter *pars* and a set of keywords **W** as the input, this encryption algorithm outputs a ciphertext CT. This algorithm is run by the data owner.
- Test(*pars*, sk_s, CT, T_M) $\rightarrow$ 1/0. Taking the public parameter *pars*, the server private key sk_s, a cipher- text CT associated with a keywords set **W** and a trapdoor T_M for an access structure (M, ρ,{$W_{\rho(i)}$ }) as the input, this testing algorithm outputs either 1 when the ciphertext satisfies the access structure of the trapdoor T_M or 0 otherwise. This algorithm is run by the designated server.

We require that a expressive keyword search scheme Π is correct, meaning that for all the sets of keyword- s **W** and access structures M such that M(**W**) = 1, if (*pars*, *msk*) $\leftarrow$ Setup(1^λ), (pk_s, sk_s) $\leftarrow$ sKeyGen(*pars*), T_M $\leftarrow$ Trapdoor(*pars*, pk_s, *msk*, (M, ρ, $W_{\rho(i)}$ }), CT } $\leftarrow$ Encrypt(*pars*, **W**), then Test(*pars*, sk_s, CT, T_M) = 1.

Following the security model introduced in [7], [25], we give the security definition for an expressive keyword search scheme over encrypted data in terms of the semantic security to ensure that such a scheme does not reveal any information about the keyword values in the ciphertext, which we call "indistinguishability against chosen keyword- set attack (IND-CKA)". Formally, we describe the IND-CKA security in the following game between a challenger algorithm B and an adversary algorithm A, where algorithm A is divided into algorithm A_1 (which is assumed to be a designated cloud (searching) server) and algorithm A_2 (which is assumed to be an outside attacker).

1) The security game between algorithm B and algorithm A_1 is to guarantee that the searching server cannot tell which ciphertext encrypts which set of keywords without obtaining the trapdoors for the access structures that can be satisfied by the keywords associated with the ciphertexts. This is because once the server is given a trapdoor that the keyword set in a ciphertext can satisfy, the server will ascertain that this ciphertext contains at least the keywords associated with the access structure in the given trapdoor.
   - Setup. Algorithm B runs the Setup algorithm to obtain the public parameter *pars* and the master private key *msk*. It gives the public parameter *pars* to algorithm A_1 and keeps *msk* to itself. In addition, algorithm B runs the sKeyGen algorithm to obtain a public and private key pair (pk_s, sk_s) for the server. It then gives (pk_s, sk_s) to algorithm A_1.
   - Phase 1. Algorithm A_1 adaptively issues queries to algorithm B for the trapdoors corresponding to the access structures (M_1, ρ_1, {$W_{\rho_1(i)}$}), ..., (M_{q_1}, ρ_{q_1}, {$W_{\rho_{q_1}(i)}$}). For each (M_j, ρ_j, {$W_{\rho(i)}$}$_j$) with $j \in [1, q_1]$, algorithm B runs the Trapdoor algorithm, and sends T_{M_j} to algorithm A_1.
   - Challenge. Algorithm A_1 outputs two sets of keyword $\mathbf{W}^*_0$, $\mathbf{W}^*_1$ of the same size with the restriction that $\mathbf{W}^*_0$ and $\mathbf{W}^*_1$ satisfy none of the queried trapdoors. Algorithm B selects a random bit $\beta \in \{0, 1\}$, runs the Encrypt algorithm on $\mathbf{W}^*_\beta$ to obtain the challenge ciphertext CT^*, and then forwards CT^* to algorithm A_1.
   - Phase 2. Algorithm A_1 continues issuing queries to algorithm B for the trapdoors corresponding to the access structures (M_{q_1+1}, ρ_{q_1+1}, { $W_{\rho_{q_1+1}(i)}$ }), ..., (M_q, ρ_q, { $W_{\rho_q(i)}$ }) with the restriction that any (M_j, ρ_j, $W_{\rho_j(i)}$ }) for $j \in [q+1, q]$ can be satisfied by neither $\mathbf{W}_0$ nor $\mathbf{W}_1$.
   - Guess. Algorithm A_1 outputs its guess $\beta' \in \{0, 1\}$ and wins the game if $\beta' = \beta$.

2) The security game between algorithm B and algorithm A_2 is to ensure that the outsider attacker who has not obtained the searching server's private key cannot determine the set of keyword values associated with the ciphertext even though the attacker gets the trapdoors over the access structures satisfied by the keywords associated with the ciphertexts. This is because the server's public key is embedded in the trapdoors such that no one can determine whether a trapdoor matches the keyword set of a ciphertext without the server's private key.

ISBN 978-93-88808-61-3

- Setup. Algorithm B runs the Setup algorithm to obtain the public parameter *pars* and the master private key *msk*. It gives the public parameter *pars* to algorithm A_2 and keeps *msk* to itself. Also, algorithm B runs the sKey- Gen algorithm to obtain a public and private key pair (pk_s, sk_s) for the server. It then gives pk_s to algorithm A_2 and keeps sk_s to itself.
- Phase 1. Algorithm A_2 adaptively issues queries to algorithm B for the trapdoors cor- responding to the access structures $(M_1, \rho_1, \{W_{\rho_1(i)}\}), \ldots, (M_{q_1}, \rho_{q_1}, \{W_{\rho_{q_1}(i)}\})$. For each M_j with $j \in [1, q_1]$, algorithm B runs the trapdoor generation algorithm Trapdoor, and sends T_{M_j} to algorithm A_2.
- Challenge. Algorithm A_2 outputs two sets of keyword $\mathbf{W}_0^*, \mathbf{W}_1^*$ of the same size. Algorithm B selects a random bit $\beta \in \{0, 1\}$, runs the Encrypt algorithm on $\mathbf{W}_\beta^*$ to obtain the challenge ciphertext CT^*, and then gives CT^* to algorithm A_2.
- Phase 2. Algorithm A_2 continues issuing queries to algorithm B for the trapdoors corresponding to the access structures $(M_{q_1+1}, \rho_{q_1+1}, \{W_{\rho_{q_1+1}(i)}\}), \ldots, (M_q, \rho_q, \{W_{\rho_q(i)}\})$.
- Guess. Algorithm A_2 outputs its guess $\beta' \in \{0, 1\}$ and wins the game if $\beta' = \beta$.

For $A \in \{A_1, A_2\}$, an expressive keyword search system Π is IND-CKA secure if the advantage function referring to the security game $\text{Game}^{\text{IND}}_{\Pi, A}$

$$\mathbf{Adv}^{\text{IND}}(\lambda) \stackrel{\text{def}}{=} \Pr[\beta = \beta'] \quad \cdot C^*, D^*, \{D^*, E^*_{1,1}, E^*_{i,2}, F^*_{i,1}, F^*_{i,2}\}$$

is negligible in the security parameter λ for any probabilistic polynomial-time (PPT) adversary algorithm A.

In addition, an expressive keyword search system is said to be selectively IND-CKA secure[9] if an Init stage is added before the Setup phase where algorithm A commits to the challenge keyword sets $\mathbf{W}_0^*, \mathbf{W}_1^*$ which it aims to attack.

APPENDIX B
SECURITY PROOF OF THEOREM 1

Proof. The proof is divided into two parts, depending on the role of the adversary. In the first part, the adversary is assumed to be an outside attacker, and in the second part, the adversary is assumed to be the server.

In terms of the first part of the proof, we prove it via a sequence of games, where game Game_0 is the same as the original game, and game Game_1 is the same as Game_0 except that the trapdoors might be generated in a different way. We finish the proof by showing that if there exists an outside adversary algorithm A that can distinguish game Game_1 from game Game_0, then we can build a challenger algorithm B that solves the decisional BDH assumption.

9. Note that selective IND-CKA security is weaker than IND-CKA security, but it is a useful tool in security reduction and is widely used in the cryptographic systems.

- Algorithm A gives algorithm B two challenge keyword sets $\mathbf{W}_0^* = \{W_{0,1}^*, \ldots, W_{0,m}^*\}$ and $\mathbf{W}_1^* = \{W_{1,1}^*, \ldots, W_{1,m}^*\}$.
- Setup. Algorithm B runs the Setup algorithm to generate the public parameter and the master private key as required, and sets the public and private key pair for the server as (g^a, a). Also, algorithm B selects a random bit $\beta \in \{0, 1\}$.
- Phase 1. Since algorithm B knows the master private key, it easily outputs the trapdoor on any access structures as required. If algorithm A issues a trap- door generation query on an access structure that can be satisfied by $\mathbf{W}_\beta^*$, algorithm B computes $T = g^c$, $T' = g^b$, $T_{i,2} = H(\hat{e}(g, g)^{abc}) g^{d_1 d_2 t_{i,1} + d_3 d_4 t_{i,2}}$, and generates the other elements of the trapdoor as in the Trapdoor algorithm.
- Challenge. Algorithm B runs the Encrypt algorithm on $\mathbf{W}_\beta^*$ to obtain the challenge ciphertext CT^*, and gives CT^* to algorithm A.
- Phase 2. The same as that in Phase 1.
- Guess. Algorithm A output a guess β' for β.

On the one hand, if $Z = \hat{e}(g, g)^{abc}$, then algorithm A's view of this simulation is identical to the original game. On the other hand, if Z is randomly chosen from tt_1, then algorithm A's advantage is nil. Therefore, if algorithm A can discern game Game_1 from game Game_0 with a non-negligible probability, algorithm B has a non-negligible probability in breaking the decisional BDH problem.

Concerning the second part of the proof, we prove the security using a sequence of games. For simplicity, we remove the access structure from the ciphertext, and denote $\{D^*, E^*_{i,1}, E^*_{i,2}, F^*_{i,1}, F^*_{i,2}\}_{i \in [1,m]}$ as the challenge ciphertext given to the adversary during an attack in the real world. Let Z be a random element of tt_1, and $\{Z_{i,1}\}$, $\{Z_{i,1}\}$ be random elements of tt. We define the following games which differ on the type of the challenge ciphertext is given by the challenger to the adversary.

- Game_0: The challenge ciphertext is $CT_0 = C^*, D^*, \{D^*, E^*_{i,1}, E^*_{i,2}, F^*_{i,1}, F^*_{i,2}\}_{i \in [1,m]}$.
- Game_1: The challenge ciphertext is $CT^* = C^*, Z, D^*, \{D^*, E^*_{i,1}, E^*_{i,2}, F^*_{i,1}, F^*_{i,2}\}_{i \in [1,m]}$.
- Game_2: The challenge ciphertext is $CT_2 = C^*, Z, D^*, (D^*, Z_{1,1}, E^*_{1,1}, E^*_{1,2}, F^*_{1,1}, F^*_{1,2}), \{D^*, E^*_{i,1}, E^*_{i,2}, F^*_{i,1}, F^*_{i,2}\}_{i \in [2,m]}$.
- ……
- Game_{m+1}: The challenge ciphertext is $CT^*_m = C^*, Z, D^*, \{D^*, Z^*_{i,1}, E^*_{i,2}, F^*_{i,1}, F^*_{i,2}\}_{i \in [1,m]}$.
- Game_{m+2}: The challenge ciphertext is $CT_{m+2} = C^*, Z, D^*, (D^*, Z_{1,1}, E^*_{1,1}, E^*_{1,2}, F^*_{1,1}, F^*_{1,2}), \{D^*, Z_{i,1}, E^*_{i,2}, F^*_{i,1}, F^*_{i,2}\}_{i \in [2,m]}$.
- ……
- Game_{2m+1}: The challenge ciphertext is $CT_{2m+1} = C^*, Z, D^*, \{D^*, Z_{i,1}, E^*_{i,2}, Z^*_{i,1}, F^*_{i,2}\}_{i \in [1,m]}$.

To complete the proof, we will show that the games Game_0, Game_1, ..., Game_{2m+1} are computationally indistinguishable from each other.

Lemma 1. Under the $(q–2)$ assumption, the advantage for a polynomial time adversary that can distinguish between the games Game_0 and Game_1 is negligible.

ISBN 978-93-88808-61-3

Proof. Assume that there is an adversary algorithm A that can distinguish $Game_0$ from $Game_1$. Then we can build a challenger algorithm B that can solve the $(q-2)$ problem.

- Init. Algorithm A gives algorithm B two challenge keyword sets $\mathbf{W}^*_0 = \{W^*_{0,1}, ..., W^*_{0,m}\}$ and $\mathbf{W}^*_1 = \{W^*_{1,1}, ..., W^*_{1,m}\}$.
- Setup. In order to generate the public system parameter, algorithm B implicitly sets $\alpha = xy$. Then algorithm B randomly chooses $\beta \in \{0, 1\}$, $d_1, d_2, d_3, d_4, \tilde{u}, \tilde{h} \in Z_p$, and computes the public parameter $pars = (g, u, h, w, g_1, g_2, g_3, g_4, \hat{e}(g^x, g^y))$ as follows.

$$g = g,\ w = g^x,\ g_1 = g^{d_1},\ g_2 = g^{d_2},$$

$$g_3 = g^{d_3},\ g_4 = g^{d_4},\ u = g^{\tilde{u}} \cdot \prod_{i\in[m]} g^{y/b_i^2},$$

$$h = g^{\tilde{h}} \cdot \prod_{i\in[m]} (g^{xz/b_i}) \cdot \prod_{i\in[m]} (g^{y/b_i^2})^{-W^*_{\beta,i}}.$$

- Phase 1 and Phase 2. Algorithm B has to create the trapdoors for the access structures $(M, \rho, \{\rho(i)\})$[10] required by algorithm A that are not satisfied by either $\mathbf{W}^*_0$ or $\mathbf{W}^*_1$. Since $\mathbf{W}^*_\beta$ does not satisfy (M, ρ), there exists a vector $\vec{w} = (w_1, ..., w_n) \in Z_p^n$ such that $w_1 = 1$ and $M_i \cdot \vec{w} = 0$ for all $i \in [l]$ such that $\rho(i) \in \mathbf{W}^*_\beta$. Algorithm B computes $\vec{w}$ using linear algebra. The vector $\vec{y}$ shared is implicitly set as $\vec{y} = xy\vec{w} + (0, \tilde{y}_2, ..., \tilde{y}_n)^\perp$, where $\tilde{y}_2, ..., \tilde{y}_n \in Z_p$. This is a properly distributed vector with first component as $xy = \alpha$ and the other components being uniformly random in Z_p. As a result, for each row $i \in [l]$, the share is

$$v_i = M_i \cdot \vec{y} = xy(M_i \cdot \vec{w}) + (M_i \cdot (0, \tilde{y}_2, ..., \tilde{y}_n)^\perp)$$

$$= xy(M_i \cdot \vec{w}) + \tilde{v}_i.$$

As mentioned above, for each row i for which $\rho(i) \in \mathbf{W}^*_\beta$, $M_i \cdot \vec{w} = 0$. In this case, $v_i = \tilde{v}_i = M_i \cdot (0, \tilde{y}_2, ..., \tilde{y}_n)^\perp$, which is known to algorithm B, so algorithm B randomly chooses $t_i \in Z_p$, and outputs $\{T_{i,1}, T_{i,2}, T_{i,3}, T_{i,4}, T_{i,5}, T_{i,6}\}$ as in the Trapdoor algorithm.

For each row $i \notin \mathbf{W}^*_\beta$, algorithm B randomly chooses $\tilde{t}_{i,1}, \tilde{t}_{i,2} \in Z_p$, and implicitly sets

$$T_{i,1} = g^{v_i} w^{d_1 d_2 t_{i,1} + d_3 d_4 t_{i,2}}$$

$$= g^{\tilde{v}_i} \cdot \Big(\prod_{j\in[n]} (g^{xzb_j})^{\frac{M_i \cdot \vec{w}}{\rho(i)-W^*_{\beta,j}}} \cdot w^{\tilde{t}_{i,1}}\Big)^{d_1 d_2} \cdot \Big(\prod_{j\in[n]} (g^{xzb_j})^{\frac{M_i \cdot \vec{w}}{\rho(i)-W^*_{\beta,j}}} \cdot w^{\tilde{t}_{i,2}}\Big)^{d_3 d_4}.$$

$$T_{i,2} = g^{d_1 d_2 t_{i,1} + d_3 d_4 t_{i,2}}$$

$$= (g^y)^{-M_i \cdot \vec{w}} \cdot \Big(\prod_{j\in[m]} (g^{xzb_j})^{\frac{M_i \cdot \vec{w}}{\rho(i)-W^*_{\beta,j}}} \cdot g^{\tilde{t}_{i,1}}\Big)^{d_1 d_2} \cdot \Big(\prod_{j\in[m]} (g^{xzb_j})^{\frac{M_i \cdot \vec{w}}{\rho(i)-W^*_{\beta,j}}} \cdot g^{\tilde{t}_{i,2}}\Big)^{d_3 d_4}.$$

$$T_{i,3} = ((u^{\rho(i)}h)^{-t_{i,1}})^{-d_2}$$

$$= (g^y)^{(M_i \cdot \vec{w})(\rho(i)\tilde{u}+\tilde{h})} \cdot \prod_{j\in[m]} (g^{xzb_j})^{\frac{\rho(i)\tilde{u}+\tilde{h}}{\rho(i)-W^*_{\beta,j}}} \cdot \prod_{(j,k)\in[m,m]} (g^{(xz)^2 b_k/b_j})^{\frac{-(M_i \cdot \vec{w})}{\rho(i)-W^*_{\beta,k}}} \cdot \prod_{j\in[m]} (g^{y^2/b_j^2})^{(M_i \cdot \vec{w})(\rho(i)-W^*_{\beta,j})} \cdot \prod_{(j,k)\in[m,m],\ j\neq k} (g^{xzb_k/b_j^2})^{\frac{-(M_i \cdot \vec{w})(\rho(i)-W^*_{\beta,j})}{\rho(i)-W^*_{\beta,k}}} \cdot ((u^{\rho(i)}h)^{-\tilde{t}_{i,1}})^{-d_2}.$$

Since $T_{i,3}$, $T_{i,4}$, $T_{i,5}$, $T_{i,6}$ have the term $(u^{\rho(i)}h)^{-t_{i,1}}$ in common, and d_1, d_2, d_3 and d_4 are known to algorithm B, algorithm B can simply compute $T_{i,4}$, $T_{i,5}$, $T_{i,6}$ as $T_{i,3}$. Thus, algorithm B successfully responds to algorithm A's trapdoor queries.

- Challenge. To generate a challenge ciphertext, algorithm B implicitly sets $\mu = z$ from the $q-2$ assumption, and $z_i = b_i$ for every $i \in [m]$. Notice that these parameters are properly distributed since $z, b_1, ..., b_q$ are information theoretically hidden from the view of algorithm A. In addition, algorithm B

ISBN 978-93-88808-61-3

On the one hand, if $Z = \hat{e}(g, g)^{xyz}$, then algorithm A's view of this simulation is identical to the original game. On the other hand, if Z is randomly chosen from tt_1, then algorithm A's advantage is nil. Therefore, if algorithm A can distinguish game $Game_1$ from game $Game_0$ with a non-negligible probability, algorithm B has a non-negligible probability in breaking the $(q-2)$ assumption.

Lemma 2. Under the decisional linear assumption, the advantage for a polynomial time adversary that can distinguish between the games $Game_{m+1}$ and $Game_m$ for $m \in [1, m]$ is negligible.

Proof. Assuming that there is an adversary algorithm A that can distinguish $Game_m$ from $Game_{m+1}$, we can build a challenger algorithm B to solve the decisional linear problem.

- Init. Algorithm A gives algorithm B two challenge keyword sets $\mathbf{W}_0 = \{W^*_{0,1}, ..., W^*_{0,m}\}$, $\mathbf{W}^* = \{W^*_{1,1}, ..., W^*_{1,m}\}$.
- Setup. In order to generate the public system parameter, algorithm B implicitly sets $d_1 = x_2$, $d_2 = x_1$. Then algorithm B randomly chooses $d_3, d_4, \beta \in \{0, 1\}$, $\alpha, y, \tilde{w} \in Z_p$, and computes the public parameter $pars_{m+m+1} = (g, u, h, w, g_1, g_2, g_3, g_4, \hat{e}(g, g)^{\alpha})$.

$$g = g, \quad w = g^{\tilde{w}}, \quad g_1 = g^{x_2}, \quad g_2 = g^{x_1},$$
$$g_3 = g^{x_3}, \; g_4 = g^{x_4}, \; u = g^{x_2\alpha},$$
$$h = g^{-x_2\alpha W^*_{\beta,m}} g^{y}, \quad \hat{e}(g, g)^{\alpha} = \hat{e}(g, g)^{\alpha}.$$

- Phase 1 and Phase 2. In order to create a trapdoor for an access structure (M, ρ) required by algorithm A that is satisfied by neither $\mathbf{W}^*_0$ nor $\mathbf{W}^*_1$, algorithm B performs as follows. It randomly chooses $\vec{y} = (\alpha, y_2, ..., y_n)^{\perp}$ where $y_2, ..., y_n \in Z_p$. Also, it randomly chooses $t_{1,1}, t_{1,2}, ..., t_{l,1}, t_{l,2} \in Z_p$. For each $i \in [l]$, algorithm B sets $v_i = M_i \cdot \vec{y}$,

$$\tilde{t}_{i,1} = \frac{t_{i,1}\alpha(\rho(i) - W^*_{\beta,m})}{x_2\alpha(\rho(i) - W^*_{\beta,m}) + y'} + \frac{y x_1 t_{i,1}}{d_3 d_4(\rho(i) - W^*_{\beta,m})x_2 + y}.$$

Then it outputs the trapdoor as

$$T_{i,1} = g^{v_i}(g^{x_1 t_{i,1}} g^{t_{i,2} d_3 d_4})^{\tilde{w}} = g^{v_i w^{d_1 d_2 \tilde{t}_{i,1} + d_3 d_4 \tilde{t}_{i,2}}},$$
$$T_{i,2} = g^{x_1 t_{i,1}} g^{t_{i,2} d_3 d_4} = g^{x_1 x_2 \tilde{t}_{i,1} + \tilde{t}_{i,2} d_3 d_4}$$
$$= g^{d_1 d_2 \tilde{t}_{i,1} + d_3 d_4 \tilde{t}_{i,2}},$$
$$T_{i,3} = (g^{x_1})^{-\alpha(\rho(i) - W^*_{\beta,m})t_{i,1}} = ((u^{\rho(i)} h)^{\tilde{t}_{i,1}})^{-d_2},$$
$$T_{i,4} = (g^{x_2})^{-\alpha(\rho(i) - W^*_{\beta,m})t_{i,1}} = ((u^{\rho(i)} h)^{\tilde{t}_{i,1}})^{-d_1},$$
$$T_{i,5} = (g^{x_1})^{-\alpha(\rho(i) - W^*_{\beta,m})t_{i,2}} = ((u^{\rho(i)} h)^{\tilde{t}_{i,2}})^{-d_2},$$
$$T_{i,6} = (g^{x_2})^{-\alpha(\rho(i) - W^*_{\beta,m})t_{i,2}} = ((u^{\rho(i)} h)^{\tilde{t}_{i,2}})^{-d_1}.$$

- Challenge. To generate a challenge ciphertext, algorithm B implicitly sets $s_{m,1} = x_3$, $z_m = x_3 + x_4$ from the decisional linear assumption. In addition, algorithm B randomly chooses $\mu, s_{1,1}, ..., s_{m-1,1}, s_{1,2}, ..., s_{m,2} \in Z_p$, $z_1, ..., z_{m-1} \in Z_p$. Thus, algorithm B can calculate the challenge ciphertext as follows.

1) For $i = m$, algorithm B outputs

$$C^* = \hat{e}(g, g)^{\alpha\mu}, \quad D^* = g^{\mu},$$
$$D_m = w^{-\mu}(u^{\beta,m} h)^{y} = w^{-\mu y} Z,$$
$$E^*_{m,1} = g_1^{z_m - s_1} = g^{x_2 x_4},$$
$$E^*_{m,2} = g_2^{s_{m,1}} = g^{x_1 x_3},$$
$$F^*_{m,1} = Z^{d_3} \cdot g_3^{-s_{m,2}}, \; F^*_{m,2} = g_4^{s_{m,2}}.$$

2) For any $i \in [m-1]$, algorithm B outputs

$$D^* = w^{-\mu}(u^{W^*_{\beta,i}} h)^{z_i}, \quad E^*_{i,1} = g_1^{z_i - s_{i,1}},$$
$$E^*_{i,2} = g_2^{s_{i,1}}, \; F^*_{i,1} = g_3^{z_i - s_{i,2}}, \; F^*_{i,2} = g_4^{s_{i,2}}.$$

- Guess. Algorithm A output a guess β' for β.

On the one hand, if $Z = g^{x_3+x_4}$, then algorithm A's view of this simulation is identical to the original game. On the other hand, if Z is randomly chosen from tt, then algorithm A's advantage is nil. Therefore, if algorithm A can distinguish game $Game_m$ from game $Game_{m+1}$ with a non-negligible probability, algorithm B has a non-negligible probability in breaking the decisional linear assumption.

Lemma 3. Under the decisional linear assumption, the advantage for a polynomial time adversary that can distinguish between the games $Game_t$ and $Game_t$ for $m' \in [1, m]$ is negligible.

proof. This proof follows almost the same as that of Lemma 2, except that the simulation is done over the parameters g_3 and g_4 instead of g_1 and g_2.

This completes the proof of Theorem 1.

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

1

A Comprehensive Study on Social Network Mental Disorders Detection via Online Social Media Mining

Naveen

M.Tech Scholar,
CSE Department
Malla Reddy College of Engineering
naveenit@gmail.com

Abstract—The explosive growth in popularity of social networking leads to the problematic usage. An increasing number of social network mental disorders (SNMDs), such as Cyber-Relationship Addiction, Information Overload, and Net Compulsion, have been recently noted. Symptoms of these mental disorders are usually observed passively today, resulting in delayed clinical intervention. In this paper, we argue that mining online social behavior provides an opportunity to actively identify SNMDs at an early stage. It is challenging to detect SNMDs because the mental status cannot be directly observed from online social activity logs. Our approach, new and innovative to the practice of SNMD detection, does not rely on self-revealing of those mental factors via questionnaires in Psychology. Instead, we propose a machine learning framework, namely, *Social Network Mental Disorder Detection (SNMDD)*, that exploits features extracted from social network data to accurately identify potential cases of SNMDs. We also exploit multi-source learning in SNMDD and propose a new SNMD-based Tensor Model (STM) to improve the accuracy. To increase the scalability of STM, we further improve the efficiency with performance guarantee. Our framework is evaluated via a user study with 3126 online soc ial network users. We conduct a feature analysis, and also apply SNMDD on large-scale datasets and analyze the characteristics of the three SNMD types. The results manifest that SNMDD is promising for identifying online social network users with potential SNMDs.

Index Terms—Tensor factorization acceleration, online social network, mental disorder detection, feature extraction.

1 INTRODUCTION

With the explosive growth in popularity of social network- ing and messaging apps, online social networks (OSNs) have become a part of many people's daily lives. Most research on social network mining focuses on discovering the knowledge behind the data for improving people's life. While OSNs seemingly expand their users' capability in increasing social contacts, they may actually decrease the face-to-face interpersonal interactions in the real world. Due to the epidemic scale of these phenomena, new terms such as Phubbing (Phone Snubbing) and Nomophobia (No Mobile Phone Phobia) have been created to describe those who cannot stop using mobile social networking apps.

- *Hong-Han Shuai is with the Department of Electrical Computer Engineering, National Chiao Tung University, No. 1001, University Road, Hsinchu, Taiwan. E-mail: hhshuai@nctu.edu.tw*
- *Chih-Ya Shen is with the Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan. E-mail: chihya@cs.nthu.edu.tw*
- *De-Nian Yang and Ming-Syan Chen are with the Research Cen- ter of Information Technology Innovation, Academia Sinica, No. 128, Sec. 2, Academia Road, Taipei, 11529 Taiwan. Email: {dnyang, mschen}@citi.sinica.edu.tw.*
- *Yi-Feng Lan is with the Graduate Institute of Educational Psychology and Counseling, Tamkang University, Taiwan. E-mail: carolyflan@gmail.com.*
- *Wang-Chien Lee is with the Department of Computer Science and Engineering, Pennsylvania State University, PA, USA. E-mail: wlee@cse.psu.edu.*
- *Philip S. Yu is with the Department of Computer Science, University of Illinois at Chicago, IL, USA. E-mail: psyu@uic.edu.*
- *Ming-Syan Chen is also with the Department of Electrical Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, 10617 Taiwan.*

In fact, some social network mental disorders (SNMDs), such as Information Overload and Net Compulsion [1], have been recently noted.[1] For example, studies point outthat 1 in 8 Americans suffer from problematic Internet use[2]. Moreover, leading journals in mental health, such as the American Journal of Psychiatry [2], have reported that the SNMDs may incur excessive use, depression, social with- drawal, and a range of other negative repercussions.

Indeed, these symptoms are important components of diagnostic criteria for SNMDs [3] e.g., excessive use of social networking apps – usually associated with a loss of the sense of time or a neglect of basic drives, and with- drawal – including feelings of anger, tension, and/or de- pression when the computer/apps are inaccessible. SNMDs are social-oriented and tend to happen to users who usually interact with others via online social media. Those with SNMDs usually lack offline interactions, and as a result seek cyber-relationships to compensate. Today, identification of potential mental disorders often falls on the shoulders of supervisors (such as teachers or parents) passively. How- ever, since there are very few notable physical risk factors, the patients usually do not actively seek medical or psycho- logical services. Therefore, patients would only seek clinical interventions when their conditions become very severe.

However, a recent study shows a strong correlation be- tween suicidal attempt and SNMDs [4], which indicates that adolescents suffering from social network addictions have a much higher risk of suicidal inclination than non-addictive

1. http://phys.org/news/2015-09-social-media-impacts-mental- well-being.html
2. http://netaddiction.com/faqs/

users. The research also reveals that social network addic- tion may negatively impact emotional status, causing higher hostility, depressive mood, and compulsive behavior. Even more alarming is that the delay of early intervention may seriously damage individuals' social functioning. In short, it is desirable to have the ability to actively detect potential SNMD users on OSNs at an early stage.

Although previous work in Psychology has identified several crucial mental factors related to SNMDs, they are mostly examined as standard diagnostic criteria in survey questionnaires. To automatically detect potential SNMD cases of OSN users, extracting these factors to assess users' online mental states is very challenging. For example, the extent of loneliness and the effect of disinhibition of OSN users are not easily observable.[3] Therefore, there is a need to develop new approaches for detecting SNMD cases of OSN users. We argue that mining the social network data of individuals as a complementary alternative to the con- ventional psychological approaches provides an excellent opportunity to *actively identify* those cases at an early stage. In this paper, we develop a machine learning framework for detecting SNMDs, which we call *Social Network Mental Disorder Detection (SNMDD)*.

Specifically, we formulate the task as a semi-supervised classification problem to detect three types of SNMDs [1]:
i) Cyber-Relationship Addiction, which shows addictive be- havior for building online relationships; ii) Net Compulsion, which shows compulsive behavior for online social gaming or gambling; and iii) Information Overload, which is related to uncontrollable surfing. By exploiting machine learning techniques with the ground truth obtained via the current diagnostic practice in Psychology [1], we extract and analyze the following crucial categories of features from OSNs: 1) social comparison, 2) social structure, 3) social diversity,
4) parasocial relationships, 5) online and offline interaction ratio, 6) social capital, 7) disinhibition, 8) self-disclosure, and 9) bursting temporal behavior. These features capture important factors or serve as proxies for SNMD detection. For example, studies manifest that users exposed to positive posts from others on Facebook with similar background are inclined to feel malicious envy and depressed due to the social comparison [36]. The depression leads users to disorder behaviors, such as information overload or net compulsion. Therefore, we first identify positive newsfeeds and then calculate the profile similarity and relation famil- iarity between friends. As another example, a parasocial relationship is an asymmetric interpersonal relationship, i.e., one party cares more about the other, but the other does not. This asymmetric relationship is related to loneliness, one of the primary mental factors pushing users with SNMDs to excessively access online social media [5]. Therefore, we extract the ratio of the number of actions to and from friends of a user as a feature. In this paper, the extracted features are carefully examined through a user study.

Furthermore, users may behave differently on different OSNs, resulting in inaccurate SNMD detection. When the data from different OSNs of a user are available, the accu- racy of the SNMDD is expected to improve by effectively integrating information from multiple sources for model training. A na¨ıve solution that concatenates the features from different networks may suffer from the curse of dimen- sionality. Accordingly, we propose an *SNMD-based Tensor Model (STM)* to deal with this multi-source learning problem in SNMDD. Advantages of our approach are: i) the novel *STM* incorporates the SNMD characteristics into the tensor model according to Tucker decomposition; and ii) the tensor factorization captures the structure, latent factors, and cor- relation of features to derive a full portrait of user behavior. We further exploit CANDECOMP/PARAFAC (CP) decom- position based *STM* and design a stochastic gradient descent algorithm, i.e., *STM-CP-SGD*, to address the efficiency and solution uniqueness issues in traditional Tucker decom- position. The convergence rate is significantly improved by the proposed second-order stochastic gradient descent algorithm, namely, *STM-CP-2SGD*. To further reduce the computation time, we design an approximation scheme of the second-order derivative, i.e., Hessian matrix, and provide a theoretical analysis.

The contributions of this paper are summarized below.

- Today online SNMDs are usually treated at a late stage. To actively identify potential SNMD cases, we propose an innovative approach, new to the current practice of SNMD detection, by mining data logs of OSN users as an early detection system.
- We develop a machine learning framework to detect SNMDs, called *Social Network Mental Disorder Detection (SNMDD)*. We also design and analyze many important features for identifying SNMDs from OSNs, such as disinhibition, parasociality, self-disclosure, etc. The proposed framework can be deployed to provide an early alert for potential patients.
- We study the *multi-source learning* problem for SNMD detection. We significantly improve the efficiency and achieve the solution uniqueness by CP decom- position, and we provide theoretical results on non- divergence. By incorporating SNMD characteristics into the tensor model, we propose *STM* to better extract the latent factors from different sources to improve the accuracy.
- We conduct a user study with 3126 users to evaluate the effectiveness of the proposed SNMDD frame- work. To the best of our knowledge, this is the first dataset crawled online for SNMD detection. Also, we apply SNMDD on large-scale real datasets, and the results reveal interesting insights on network structures in SNMD types, which can be of interest to social scientists and psychologists.

The rest of this paper is organized as follows. Section 2 surveys the related work. Section 3 presents *SNMDD*, fo- cusing on feature extraction. Section 4 presents the proposed *STM* for multi-source learning and the acceleration of tensor decomposition with the theoretical results. Section 5 reports a user study, various analyses, and the experimental results. Finally, Section 6 concludes this paper.

3. The online disinhibition effect is a loosening (or complete aban- donment) of social restrictions and inhibitions that would otherwise be present in normal face-to-face interaction during interactions with others on the Internet.

2 Related Work

Internet Addiction Disorder (IAD) is a type of behavior addiction with the patients addicted to the Internet, just like those addicting to drugs or alcohol [3]. Many research works in Psychology and Psychiatry have studied the important factors, possible consequences, and correlations of IAD [10], [40], [41], [42]. King et al. [40] investigate the problem of sim- ulated gambling via digital and social media to analyze the correlation of different factors, e.g., grade, ethnicity. Baumer et al. [10] report the Internet user behavior to investigate the reason of addiction. Li et al. [41] examine the risk factors related to Internet addiction. Kim et al. [42] investigate the association of sleep quality and suicide attempt of Internet addicts. On the other hand, recent research in Psychology and Sociology reports a number of mental factors related to social network mental disorders. Research indicates that young people with narcissistic tendencies and shyness are particularly vulnerable to addiction with OSNs [6], [7]. However, the above research explores various negative im- pacts and discusses potential reasons for Internet addiction. By contrast, this paper proposes to automatically identify SNMD patients at the early stage according to their OSN data with a novel tensor model that efficiently integrate heterogeneous data from different OSNs.

Research on mental disorders in online social networks receives increasing attention recently [43], [44], [45]. Among them, content-based textual features are extracted from user-generated information (such as blog, social media) for senti- ment analysis and topic detection. Chang et. al [43] employ an NLP-based approach to collect and extract linguistic and content-based features from online social media to iden- tify Borderline Personality Disorder and Bipolar Disorder patients. Saha et al. [44] extract the topical and linguistic features from online social media for depression patients to analyze their patterns. Choudhury et al. [45] analyze emo- tion and linguistic styles of social media data for Major De- pressive Disorder (MDD). However, most previous research focuses on individual behaviors and their generated textual contents but do not carefully examine the structure of social networks and potential Psychological features. Moreover, the developed schemes are not designed to handle the sparse data from multiple OSNs. In contrast, we propose a new multi-source machine learning approach, i.e., STM, to extract proxy features in Psychology for different diseases that require careful examination of the OSN topologies, such as Cyber-Relationship Addiction and Net Compulsion.

Our framework is built upon support vector machine, which has been widely used to analyze OSNs in many areas [11], [12]. In addition, we present a new tensor model that not only incorporates the domain knowledge but also well estimates the missing data and avoids noise to properly handle multi-source data. Caballero et al. [8] estimate the probability of mortality in ICU by modeling the probability of mortality as a latent state evolving over time. Zhao et al. [9] propose a hierarchical learning method for event de- tection and forecasting by first extracting the features from different data sources and then learning via geographical multi-level model. However, the SNMD data from different OSNs may be incomplete due to the heterogeneity. For example, the profiles of users may be empty due to the privacy issue, different functions on different OSNs (e.g., game, check-in, event), etc. We propose a novel tensor-based approach to address the issues of using heterogeneous data and incorporate domain knowledge in SNMD detection.

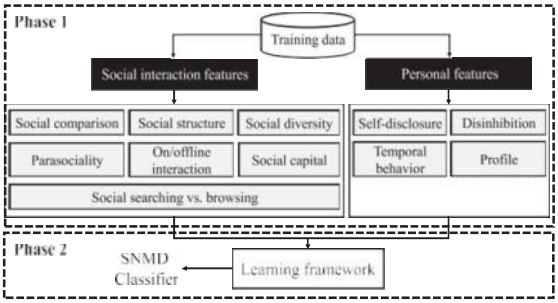


Fig. 1. The SNMDD framework.

3 Social Network Mental Disorder Detection

In this paper, we aim to explore data mining techniques to detect three types of SNMDs [1]: 1) *Cyber-Relationship (CR) Addiction*, which includes the addiction to social net- working, checking and messaging to the point where social relationships to virtual and online friends become more important than real-life ones with friends and families; 2) *Net Compulsion (NC)*, which includes compulsive online social gaming or gambling, often resulting in financial and job-related problems; and 3) *Information Overload (IO)*, which includes addictive surfing of user status and news feeds, leading to lower work productivity and fewer social inter- actions with families and friends offline.

Accordingly, we formulate the detection of SNMD cases as a classification problem. We detect each type of SNMDs with a binary SVM. In this study, we propose a two-phase framework, called *Social Network Mental Disorder Detection (SNMDD)*, as shown in Figure 1. The first phase extracts various discriminative features of users, while the second phase presents a new SNMD-based tensor model to derive latent factors for training and use of classifiers built upon Transductive SVM (TSVM) [13]. Two key challenges exist in design of SNMDD: i) we are not able to directly extract mental factors like what have been done via questionnaires in Psychology and thus need new features for learning the classification models;[4] ii) we aim to exploit user data logs from multiple OSNs and thus need new techniques for integrating multi-source data based on SNMD characteris- tics. We address these two challenges in Sections 3.1 and 4, respectively.

Feature Extraction

We first focus on extracting discriminative and informative features for design of SNMDD. This task is nontrivial for the following three reasons.

1. **Lack of mental features.** Psychological studies have shown that many mental factors are related to SNMDs, e.g., low self-esteem [3], loneliness [14]. Thus, questionnaires are designed to reveal those factors for SNMD detection. Some

4. Additional issues in feature extraction will be detailed later.

parts of Psychology questionnaire for SNMDs are based on the subjective comparison of mental states in online and offline status, which cannot be observed from OSN logs. For example:

Q1. How often do you feel depressed, moody, or nervous when you are off-line, which goes away once you are back online?
Q2. How often do you prefer the excitement of the Internet to intimacy with your partner?

Consider Q1. The feel of depression and nervousness offline can not be observed online. To tackle this problem, we have to leverage the knowledge from Psychology, such as withdrawal or relapse patterns, and exploit some proxy features extracted from online social activity logs to approxi- mate them. For Q2, the preference of excitement of the Inter- net to intimacy with users' partners is an important question for SNMD detection. As it is difficult to directly observe these factors from data collected from OSNs, psychiatrists are not able to directly assess the mental states of OSN users under the context of online SNMD detection.

2. **Heavy users vs. addictive users.** To detect SNMDs, an intuitive idea is to simply extract the usage (time) of a user as a feature for training SNMDD. However, this feature is not sufficient because i) the status of a user may be shown as "online" if she does not log out or close the social network applications on mobile phones, and ii) heavy users and addictive users all stay online for a long period, but heavy users do not show symptoms of anxiety or depression when they are not using social apps. How to distinguish them by extracting discriminative features is critical.

3. **Multi-source learning with the SNMD characteristics.** As we intend to exploit user data from different OSNs in SNMDD, how to extract complementary features to draw a full portrait of users while considering the SNMD characteristics into the tensor model is a challenging problem.

To address the first two challenges, we identify a number of effective features as proxies to capture the mental states of users, e.g., self-esteem [3] and loneliness [14].[5] The goal is to distinguish users with SNMDs from normal users. Two types of features are extracted to capture the social inter- action behavior and personal profile of a user. Due to the space constraint, some of the above features are presented in Appendix A. It is worth noting that each individual feature cannot precisely classify all cases, as research shows that exceptions may occur. Therefore, it is necessary to exploit multiple features to effectively remove exceptions.

Social Interaction Features

We first extract a number of *social interaction features* to capture the user behavior on social media.

Social comparison based features (SComp) Although most literature indicates that the majority of the newsfeed up- dates is positive, recent studies manifest that users who are exposed to positive posts from others on Facebook are in- clined to feel envy and depressed due to social comparison [38]. The social comparison leads to SNMDs according to Festinger's theory, which states that many people usually have a strong motivation to evaluate their own opinions and abilities by implicitly or explicitly comparing with others in similar backgrounds, especially when the reference in comparison to the physical world is not specific. The situa- tion becomes increasingly serious because status exchanges among friends are now very convenient via various online social networks.

Envy usually appears after comparisons, and two kinds of envy, i.e., benign envy and malicious envy, exist in Psychology [36]. The experience of benign envy leads to a moving-up motivation aiming at improving one's own position, whereas the experience of malicious envy pro- duces a pulling-down motivation and depression. Malicious envy is incurred from the comparison among close friends with similar backgrounds and states, and it usually leads to SNMDs, such as information overload or net compul- sion, because a person in this case usually feels pressure and tends to frequently check the updated status of the corresponding friends. A teenager student in this case may seek online games or gambles as alternatives for acquiring the sense of accomplishment. By contrast, benign envy is usually generated from distant friends with different back- grounds and rarely leads to SNMDs.

Therefore, for malicious envy, we first exploit the ex- isting techniques of emotional signal processing [17] to identify positive newsfeeds and then calculate the profile similarity and relation familiarity between friends. Specifi- cally, let $N_p(i, j)$ and $s(i, j)$ denote the number of positive newsfeeds that user j receives from i and the similarity on backgrounds between user i and j, respectively. For user j, the weighted number of positive newsfeeds based on similarity can be derived as

$$\frac{\sum_{i \in N(j)} [s(i, j) N_p(i,j)]}{\sum_{i \in N(j)} s(i, j)}, \quad (1)$$

where $N(j)$ is the set of neighbors of user j. Moreover, the weighted numbers of positive newsfeeds based on familiar- ity can also be derived in a similar manner by substituting the similarity function with the familiarity function as an additional proxy feature for the social comparison.

Social structure based features (SS) In Sociology, each person in a social network belongs to one of the following three types of social roles: influential users, structural holes, and normal users. An influential user is the one with a huge degree and many mentions and shares (retweets) [28]. On the other hand, weaker connecting paths between groups are structure holes in OSNs, and researchers have demon- strated that structural holes usually have timely access to important information, e.g., trade trend, job opportunities, which usually leads to social success. Therefore, the users with their roles as structural holes are more inclined to suf- fer from information overload for newsfeeds because they enjoy finding and sharing new and interesting information to various friends.

According to the above observations, we exploit the state-of-the-art approach [27] to quantify users' tendencies of being structural holes. Specifically, given n users and m communities, let $F \in R^{n \times m}$ denote the community indicator matrix, where $f_{ij} = 1$ if a user i is assigned to the j-th community, and 0 otherwise. Let $\mathbf{f}^i$ denote the i-th row vector of F. By embedding the harmonic function to learn

5. The third challenge is addressed in Section 4.

ISBN 978-93-88808-61-3

the community indicator matrix, the difference between the value of f^i and the averaged value of its neighbors $\frac{1}{d_i}\sum_{(v_i,v_j)\in E} f^j$ is required to be minimized, because neighbors are usually within similar communities. Therefore, the following minimization problem is formulated to detect structural hole spanners.

$$\min_F \|F - D^{-1}AF\|_{2,1} \quad (2)$$

$$\text{s.t. } F^T F = I_m, \quad (3)$$

where A and D are respectively the adjacency and degree matrices, and $\|X\|_{2,1}$ is the $l_{2,1}$ norm of X, which is the sum of the Euclidean norms of the columns of the matrix X. The structural hole spanners correspond to the ones with small $\|f^i\|$. Compared with social capital based features, the structural hole feature considers the community structure (global), while social capital features only examine the ego networks (local). On the other hand, we also extract the network topology based features, i.e., closeness centrality, betweenness centrality, eigenvector centrality, information centrality, flow betweenness, the rush index, as social structure based features for detecting SNMDs. For example, flow betweenness indicates how much information has been propagated through the node, which relates to information overload. Moreover, eigenvector centrality is a measure of the influence of a node in a network, and the score is similar to the pagerank, i.e., connections to high-scoring neighbors are inclined to increase the score of a node. Therefore, the scores of unpopular users are usually small and correlated to Cyber-Relationship Addiction.

Social diversity based features (SDiv) Researchers have observed that diversity improves the depth of people thinking for both majority or minority [35]. For example, a person with a more diverse background and many friends is less inclined to suffer from SNMDs because she is often supported by friends and thereby rarely feels lonely and isolated (two important factors correlated to SNMDs) [34]. Therefore, the impact of social network diversity is increasingly important and inspires us to incorporate them for effective SNMD detection. Specifically, the diversities of nationality, racial, ethical, religious, and education can be extracted as social diversity based features with Shannon index H as the diversity index, i.e.,

$$H = -\sum_{i=1}^{N_t} p_i \ln p_i, \quad (4)$$

where p_i and N_t are the proportion of users' friends belonging to the i-th type of attributes and the total number of types, respectively. The value H increases when the number of types N_t grows. Moreover, Shannon diversity index also increases when there is a more significant *evenness*. In other words, the diversity index is maximized when all type of attributes are of the equal quantities.

Parasocial relationship (PR). Research shows that the mental factor of loneliness is one of the primary reasons whythe users with SNMDs excessively access online social media [5]. As the loneliness of an OSN user is hard to measure, we exploit the parasocial relationship, an asymmetric interpersonal relationship between two people where one party cares more about the other but the other does not, to capture loneliness (as studies show that they are correlated [15]). The feature of parasocial relationship is represented as $|a_{out}|/|a_{in}|$, where $|a_{out}|$ and $|a_{in}|$ denote the number of actions a user takes to friends and the number of actions friends take to the user, respectively.[6] As the ratio increases, the extent of parasocial relationship also grows.

Due to the space constraint, other social interaction features are presented in Appendix A.1.

Personal Features

Temporal behavior features (TEMP). *Relapse* is the state that a person is inclined to quickly revert back to the excessive usage of social media after an abstinence period, while *tolerance* is the state that the time spent by a person with SNMDs tends to increase due to the mood modification effect.[7] It is worth noting that the above two mental states have been exploited to evaluate clinical addictions [14]. We aim to use them to distinguish *heavy users* and *addictive users* because heavy users do not suffer from relapse and tolerance in use of OSNs. An issue arising here is how to assess relapse and tolerance quantitatively.

It is observed that the use of social media by an SNMD patient is usually in the form of *intermittent bursts* [3]. Therefore, given a stream of a user's activities on an OSN, e.g., "likes", "comments", "posts", we exploit Kleinberg's burst detection algorithm [16], which is based on an infinite Markov model, to detect periods of the user's activities as bursty and non-bursty periods. The bursty period refers to a period during which the activities significantly increase. A bursty period is modeled as a bursty state q_1 in the Markov model, while a non-bursty period is correspondingly modeled as a normal state q_0. The burst detection algorithm finds a state transition sequence q for each user to divide the corresponding log (stream of activities) into bursty and non-bursty periods. Specifically, let $x = (x_1, x_2, \ldots, x_n)$ denote a sequence of n time intervals between $n+1$ consecutive activities, with theintervals distributed according to a density function, such as $f_{i_t}(x_t) = \alpha_{i_t} e^{-\alpha_{i_t} x_t}$, where α_{i_t} is either α_0 or α_1, and α_0 and α_1 are parameters that correspond to the normal and burst states, respectively, $\alpha_1 > \alpha_0$. A time interval x_t is in a burst state q_1 if $f_0(x_t) < f_1(x_t)$. Otherwise, it is in a normal state q_0. However, simply deciding the state sequence q based on this criteria results in numerous small periods. Therefore, a cost $\tau(q_i, q_j)$ is associated with a state transition from q_i to q_j to filter out noises and to ensure that each bursty or non-bursty period is sufficiently long. Therefore, the remaining issue is to find an optimal state-transition sequence q to minimize the following cost function [16],

$$c(q|x) = \sum_{t=1}^{n-1} \tau(q_i, q_{i+1}) + \sum_{t=1}^{n} (-\ln f_{i_t}(x_t)),$$

where $\tau(q_i, q_{i+1}) = 0$ if the state q_i and q_{i+1} is the same. $\tau(q_i, q_{i+1})$ is $\gamma \ln n$ otherwise, where γ is an algorithm parameter larger than 0. Notice that the state sequence that minimizes the cost depends on 1) how easy it is to jump from one state to another and 2) how well it is to comply to the rates of arrivals. After identifying the bursts, we

6. The actions include like, comment, and post in our work.

7. A patient may need to spend more time on social media to reach the happiness/excitement than before.

measure their intensity (the number of activities within a burst) and length (the time period of a burst) as the proxy features for *relapse* and *tolerance*, respectively. The (average, median, standard deviation, maximum, minimum) of both the burst intensity and burst length are included in our feature set, because they capture the characteristic of bursts. For instance, the standard deviation of the burst length for SNMD patients is usually larger than that for heavy users since heavy users constantly use OSNs, whereas the users with SNMDs increase the usage time due to tolerance.

Due to the space constraint, other personal features are presented in Appendix A.2.

4 Multi-Source Learning with Tensor Decomposition Acceleration

Many users are inclined to use different OSNs, and it is expected that data logs of these OSNs could provide enriched and complementary information about the user behavior. Thus, we aim to explore multiple data sources (i.e., OSNs) in SNMDD, in order to derive a more complete portrait of users' behavior and effectively deal with the data sparsity problem. To exploit multi-source learning in SNMDD, one simple way is to directly concatenate the features of each person derived from different OSNs as a huge vector. However, the above approach tends to miss the correlation of a feature in different OSNs and introduce interference. Thus, we explore tensor techniques which have been used increasingly to model multiple data sources be- cause a tensor can naturally represent multi-source data. We aim to employ tensor decomposition to extract common latent factors from different sources and objects. Based on tensor decomposition on $\mathcal{T}$, we present a *SNMD-based Tensor Model (STM)* in previous work [47], which enables U to incorporate important characteristics of SNMDs, such as the correlation of the same SNMD sharing among close friends.[8] Finally, equipped with the new tensor model, we conduct semi-supervised learning to classify each user by exploiting Transductive Support Vector Machines (TSVM) in Appendix B. In the following, the problem definition, notation explanation, and brief introduction are first presented for better reading.

Problem Definition and Notation Explanation

Given D SNMD features of N users extracted from M OSN sources, we construct a three-mode tensor $\mathcal{T} \in \mathbb{R}^{N \times D \times M}$, where each element $t_{ijk} \in \mathcal{T}$ represents the j-th feature of user i in source k. The objective here is to extract the latent features for each user with tensor composition from $\mathcal{T}$. Here scalars are denoted by lowercase letters, e.g., u, while vectors are denoted by boldface lowercase letters, e.g., $\mathbf{u}$. Matrices are represented by boldface capital letters, e.g., $\mathbf{U}$, and tensors are denoted by calligraphic letters, e.g., $\mathcal{T}$. The i-th row and the j-th column of a two-dimensional matrix $\mathbf{U}$ are respectively denoted by $\mathbf{u}_{i:}$ and $\mathbf{u}_{:j}$.

Tucker decomposition and CANDECOMP/PARAFAC (CP) decomposition have been widely used for extracting the latent features. In the following, we first briefly intro- duce Tucker decomposition.

8. Note that D does not capture the social correlations among friends.

Tucker Decomposition

Tucker decomposition [46] of a tensor $\mathcal{T} \in \mathbb{R}^{N \times D \times M}$ is defined as follows.

$$\mathcal{T} = \mathcal{C} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W}, \tag{5}$$

where $\mathbf{U} \in \mathbb{R}^{N \times R}$, $\mathbf{V} \in \mathbb{R}^{D \times S}$ and $\mathbf{W} \in \mathbb{R}^{M \times T}$ are latent matrices. R, S, and T are parameters to be set according to different criteria [46]. The 1-mode product of $\mathcal{C} \in \mathbb{R}^{R \times S \times T}$ and $\mathbf{U} \in \mathbb{R}^{N \times R}$, denoted by $\mathcal{C} \times_1 \mathbf{U}$, is a matrix with size $R \times S \times T$, where each element $(\mathcal{C} \times_1 \mathbf{U})_{nst} = \sum_{r=1}^{R} c_{rst} u_{rn}$. Given the input tensor matrix $\mathcal{T}$ that consists of the features of all users from every OSN, Tucker decomposition derives $\mathcal{C}$, U, V, and W to meet the above equality on $\mathcal{T}_{ndm}$ for every n, d, and m, where $\mathcal{C}$ needs to be diagonal, and U, V, and W are required to be orthogonal [46]. Matrix U effectively estimates a deficit feature (e.g., a missing feature value unavailable due to privacy setting) of an OSN from the corresponding feature of other OSNs, together with the features of other users with the similar behavior.

CP Decomposition

Although Tucker decomposition is flexible and general, it is difficult to interpret the latent features intuitively from the decomposed matrices due to complicated interactions among them. Also, ensuring identifiability is fundamental and important for tensor decomposition. Moreover, the model parameters are encouraged to be uniquely recov- ered given the observed statistics, i.e., the decomposition yields a unique solution. For Tucker decomposition, the identifiability needs to satisfy complicated criteria, e.g., the structured sparsity and symmetry constraints on the core tensor, and sparsity constraints on the inverse factors of the tensor decomposition [48]. In contrast, the latent features obtained by CANDECOMP/PARAFAC (CP) decomposition [33] are much easier to interpret due to the rank-1 compo- nent factorization of CP and its intrinsic axis property from parallel proportional profiles. Moreover, Kruskal criterion on the rank of tensors provides a sufficient condition of the identifiability. Most importantly, its computational com- plexity is much lower than Tucker decomposition, thereby allowing us to analyze SNMDs for large-scale OSNs.

Specifically, CANDECOMP/PARAFAC (CP) decomposi- tion of a tensor $\mathcal{T} \in \mathbb{R}^{N \times D \times M}$ is defined as follows.

$$\sum_{r=1}^{R} \mathbf{U}_{:r} \circ \mathbf{V}_{:r} \circ \mathbf{W}_{:r} \approx \mathcal{T}, \tag{6}$$

where $\circ$ denotes the vector outer product, and R is a positive integer representing the dimensionality of U, V, and W, i.e., $\mathbf{U}_{:r} \in \mathbb{R}^{N}$, $\mathbf{V}_{:r} \in \mathbb{R}^{D}$, and $\mathbf{W}_{:r} \in \mathbb{R}^{m}$, for $r = 1, \ldots, R$. The space of variables in CP decomposition is comprised of the elements of U, V, and W. The inner product of third-order tensors $\mathcal{X}$ and $\mathcal{Y}$ is defined as $\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_i \sum_j \sum_k X_{ijk} Y_{ijk}$. The objective function of CP decomposition is to find U, V, and W such that the decomposition is close to $\mathcal{T}$ (i.e., the difference is minimized). Each element $\mathbf{U}_{:r} \circ \mathbf{V}_{:r} \circ \mathbf{W}_{:r}$ is a rank-one tensor, and $\mathbf{U}_{i:}$ represents the SNMD feature tensor of user i. Compared to Tucker decomposition, the core tensor $\mathcal{C}$ in CP decomposition has been simplified, and thus the number of parameters required to be estimated

ISBN 978-93-88808-61-3

in Equation (6) is much smaller. Moreover, the solution is unique in CP decomposition but not unique in Tucker decomposition. Equipped with CP decomposition, the objective function L(T , U, V, W) is

$$\frac{1}{2}\|\mathcal{T}-\sum_{r=1}^{R}\mathbf{U}_{:r}\circ\mathbf{V}_{:r}\circ\mathbf{W}_{:r}\|^2+\frac{\lambda_1}{2}tr(\mathbf{U}^\top\mathbf{L}_a\mathbf{U})+\frac{\lambda_2}{2}\|\mathbf{U}\|^2, \quad (7)$$

where $tr(\cdot)$ denotes the matrix trace, the Frobenius norm of a tensor $\mathcal{T}$ is defined as $\|\mathcal{T}\|=\sqrt{<\mathcal{T},\mathcal{T}>}$, and λ_1 and λ_2 are parameters controlling the contribution of each part during the above collaborative factorization. The Laplacian matrix $\mathbf{L}_a$ of the weighted adjacency matrix A is defined as $\mathbf{D}-\mathbf{A}$, where D is a diagonal matrix with the entries $d_{ii}=\sum_j a_{ij}$. L first minimizes the decomposition error, i.e., $\|\mathcal{T}-\sum_{r=1}^{R}\mathbf{U}_{:r}\circ\mathbf{V}_{:r}\circ\mathbf{W}_{:r}\|^2$ for $\mathcal{T}$. Moreover, the term that minimizes $\|\mathbf{U}\|^2$ is to derive a more concise latent feature matrix and avoid overfitting. The proposed *STM* is different from the conventional tensor models in the second term of Eq. (7), where important characteristics of SNMDs are incorporated. For example, the probability of finding CR cases around a CR patient is higher than that around a non- CR user due to the loneliness propagation [15]. That is, CR users usually feel lonely and are more likely to establish friendships in cyberspace with other users with similar behavior. Since the nearby nodes with a great quantity of interactions tend to be the same (either CR or non-CR), it is envisaged that the distance between $\mathbf{u}_{i:}$ and $\mathbf{u}_{j:}$ will be small if the edge weight of the edge connecting user i and user j, i.e., $a_{i,j}$ in the adjacency matrix A, is sufficiently large. Therefore, a regularization (smoothing) term $\frac{\lambda_1}{2}tr(\mathbf{U}^T\mathbf{L}_a\mathbf{U})$ is included in the model to achieve the above goal. Due to the space constraint, the details of deriving $\frac{\lambda_1}{2}tr(\mathbf{U}^T\mathbf{L}_a\mathbf{U})$ are presented in Appendix C.

Stochastic Gradient-Descent Algorithm

Notice that CP decomposition is non-convex. Fortraditional gradient descent algorithms [25], the learning step size η and the initial values on U, V, and W are very sensi- tive and need to be carefully determined. Otherwise, the algorithm is inclined to diverge, consequently failing to find the decomposition solution. To address this issue, we design a new stochastic gradient-descent algorithm with low computational complexity to guarantee the solution convergence.

We present a stochastic gradient-descent algorithm for CP decomposition of the SNMD-based Tensor Model, namely, *SGD-CP-STM*, to iteratively improve each element in the matrices according to the corresponding gradient. Specifically, let $\mathcal{T}(\cdot,\mathbf{V},\mathbf{W})$ be a matrix obtained from by contracting V and W, i.e.,

$$\mathcal{T}(\cdot,\mathbf{V},\mathbf{W})_{ir}=\sum_j\sum_k\mathcal{T}_{ijk}\mathbf{V}_{jr}\mathbf{W}_{kr}, \quad (8)$$

where $\mathcal{T}(\cdot,\mathbf{V},\mathbf{W})\in\mathbb{R}^{N\times R}$ (the same as U). The following lemma first derives the gradient of each iteration.

Lemma 1. *The gradient of* L *with regard to* U, *i.e.,* $\nabla_{\mathbf{U}}\mathcal{L}(\mathcal{T},\mathbf{U},\mathbf{V},\mathbf{W})$, *is equal to*

$$-\mathcal{T}(\cdot,\mathbf{V},\mathbf{W})+\mathbf{U}(\Gamma(\mathbf{V},\mathbf{W})+\lambda_2\mathbf{I}_R)+\lambda_1\mathbf{L}_a\mathbf{U},$$

where $\Gamma(\mathbf{V},\mathbf{W})$ *is the Hadamard product of* $\mathbf{V}^\top\mathbf{V}$ *and* $\mathbf{W}^\top\mathbf{W}$, *i.e.,* $\Gamma(\mathbf{V},\mathbf{W})_{ij}=(\mathbf{V}^\top\mathbf{V})_{ij}(\mathbf{W}^\top\mathbf{W})_{ij}$, *and* $\mathbf{I}_R$ *is the identity matrix of size R.*

Proof. The objective function $\mathcal{L}(\mathcal{T},\mathbf{U},\mathbf{V},\mathbf{W})$ is comprised of three terms, and the derivative of $\frac{\lambda_2}{2}\|\mathbf{U}\|^2$ with regard to U is $\lambda_2\mathbf{I}_R$. For the first term, the CP gradient can be solved by the following equation according to [26].

$$\nabla_{\mathbf{U}}\frac{1}{2}\|\mathcal{T}-\sum_{r=1}^{R}\mathbf{U}_{:r}\circ\mathbf{V}_{:r}\circ\mathbf{W}_{:r}\|^2=-\mathcal{T}(\cdot,\mathbf{V},\mathbf{W})+\mathbf{U}\Gamma(\mathbf{V},\mathbf{W}). \quad (9)$$

For the second term, i.e., $\frac{\lambda_1}{2}tr(\mathbf{U}^\top\mathbf{L}_a\mathbf{U})$, the gradient for U is

$$\nabla_{\mathbf{U}}\frac{\lambda_1}{2}tr(\mathbf{U}^\top\mathbf{L}_a\mathbf{U})=\frac{\lambda_1}{2}(\mathbf{L}_a+\mathbf{L}_a^\top)\mathbf{U}. \quad (10)$$

If the weighted adjacency matrix A is symmetric, Equation (10) can be further simplified to $\lambda_1\mathbf{L}_a\mathbf{U}$, and $\nabla_{\mathbf{U}}\mathcal{L}(\mathcal{T},\mathbf{U},\mathbf{V},\mathbf{W})$ is equal to

$$-\mathcal{T}(\cdot,\mathbf{V},\mathbf{W})+\mathbf{U}(\Gamma(\mathbf{V},\mathbf{W})+\lambda_2\mathbf{I}_R)+\lambda_1\mathbf{L}_a\mathbf{U}. \quad (11)$$

The theorem follows. □

Therefore, the stochastic gradient descent algorithm up- dates U at the t-th iteration as follows.

$$\mathbf{U}^{(t)}=\mathbf{U}^{(t-1)}-\eta^{(t)}(-\mathcal{T}^{(t-1)}(\cdot,\mathbf{V}^{(t-1)},\mathbf{W}^{(t-1)})+\mathbf{U}^{(t-1)}(\Gamma(\mathbf{V}^{(t-1)},\mathbf{W}^{(t-1)})+\lambda_2\mathbf{I}_R)+\lambda_1\mathbf{L}_a\mathbf{U}^{(t-1)}). \quad (12)$$

Based on Eq. (9), the gradient for V and W can be derived in the similar way as follows:

$$\nabla_{\mathbf{V}}\mathcal{L}(\mathcal{T},\mathbf{U},\mathbf{V},\mathbf{W})=-\mathcal{T}(\mathbf{U},\cdot,\mathbf{W})+\mathbf{V}\Gamma(\mathbf{U},\mathbf{W})$$
$$\nabla_{\mathbf{W}}\mathcal{L}(\mathcal{T},\mathbf{U},\mathbf{V},\mathbf{W})=-\mathcal{T}(\mathbf{U},\mathbf{V},\cdot)+\mathbf{W}\Gamma(\mathbf{U},\mathbf{V}).$$

Note that $\mathbf{V}^{(t)}$ and $\mathbf{W}^{(t)}$ are also updated similarly in each iteration.

Acceleration of Convergence Rate

It has been widely recognized that the performance is poor when a constant learning step size η is adopted. If the learn- ing step size is too large, it is inclined to skip the optimal solution. In contrast, if the learning step size is too small, the convergence rate to the optimal solution becomes unac- ceptably slow. To avoid the above issue, we further propose a second-order stochastic gradient descent algorithm for CP decomposition, namely, *STM-CP-2SGD*. The second-order stochastic gradient descent (2SGD) considers the second- order information by adaptively assigning the learning rate as the inverse of the Hessian matrix in objective function L to guide the searching direction. Since the inversion of the full Hessian matrix

$$\begin{bmatrix}\nabla_{\mathbf{U}}\nabla_{\mathbf{U}}\mathcal{L} & \nabla_{\mathbf{U}}\nabla_{\mathbf{V}}\mathcal{L} & \nabla_{\mathbf{U}}\nabla_{\mathbf{W}}\mathcal{L}\\ \nabla_{\mathbf{W}}\nabla_{\mathbf{U}}\mathcal{L} & \nabla_{\mathbf{W}}\nabla_{\mathbf{V}}\mathcal{L} & \nabla_{\mathbf{W}}\nabla_{\mathbf{W}}\mathcal{L}\end{bmatrix}$$

is computationally expensive, its block-diagonal parts are alternatively used as an approximation of the inverse of the Hessian matrix [29]. Therefore, we update $\mathbf{U}^{(t)}$ as follows.

$$\mathbf{U}^{(t)}=\mathbf{U}^{(t-1)}-\eta^{(t)}(\nabla^2_{\mathbf{U}^{(t-1)}}\mathcal{L})^{-1}\nabla_{\mathbf{U}^{(t-1)}}\mathcal{L}. \quad (13)$$

ISBN 978-93-88808-61-3

To calculate $\nabla^2_U L$, we take a derivative of Eq. (11) with respect to U and obtain

$$\nabla^2_U L(T, U, V, W) = \frac{d(U(\Gamma(V, W)+\lambda_2 I_R))}{dU} + \lambda_1 \frac{d(L_a U)}{dU} \quad (14)$$

We find the derivatives by exploiting the relationship between the Kronecker product and the vec operator (vec- torizing matrices by stacking its columns) as follows. For the first term in Eq. (14), we have

$$\frac{d(U(\Gamma(V, W)+\lambda_2 I_R))}{dU} = \frac{d(\mathrm{vec}(U(\Gamma(V, W)+\lambda_2 I_R))}{d\mathrm{vec}(U)}. \quad (15)$$

Since $\mathrm{vec}(U(\Gamma(V, W) + \lambda_2 I_R)) = \mathrm{vec}(I_N U(\Gamma(V, W) + \lambda_2 I_R)) = (\Gamma(V, W) + \lambda_2 I_R)^T \otimes I_N \mathrm{vec}(U)$, where $\otimes$ is Kronecker product, we have

$$\frac{d(U(\Gamma(V, W)+\lambda_2 I_R))}{dU} = (\Gamma(V, W)+\lambda_2 I_R)^T \otimes I_N. \quad (16)$$

We find the second term in Eq. (14) in a similar manner.

$$\frac{d(L_a U)}{dU} = \frac{d\mathrm{vec}(L_a U)}{d\mathrm{vec}(U)}. \quad (17)$$

Since $\mathrm{vec}(L_a U) = \mathrm{vec}(L_a U I_R) = I_R \otimes L_a \mathrm{vec}(U)$, we have

$$\frac{d(L_a U)}{dU} = I_R \otimes L_a. \quad (18)$$

Substituting Eqs. (16) and (18) into Eq. (14), $\nabla^2_U L(T, U, V, W)$ is equal to

$$(\Gamma(V, W)+\lambda_2 I_R)^T \otimes I_N + \lambda_2 I_R \otimes L_a = (\Gamma(V, W)+\lambda_2 I_R)^T \oplus \lambda_2 L_a, \quad (19)$$

where $\oplus$ is the Kronecker sum. Therefore, weupdate U at the t-th iteration as follows, i.e., $U^{(t)}$ is equal to

$$U^{(t-1)} - \eta^{(t)}((\Gamma(V^{(t-1)}, W^{(t-1)})+\lambda_2 I_R)^T \oplus \lambda_2 L_a)^{-1} \nabla_{U^{(t-1)}} L. \quad (20)$$

Compared with Eq. (12), here the update of U includes a new term, i.e., $((\Gamma(V, W)+\lambda_2 I_R)^T \oplus \lambda_2 L_a)^{-1}$, representing the adaptive learning step size. Previous studies show that the number of iterations for 2SGD to reach the optimum is much smaller than that of SGD [49]. More specifically, given a second-order convex function, 2SGD requires only one iteration because it derives the optimal step size to the opti- mum point. To properly choose the $\eta^{(t)}$, several approaches can be applied,e.g., damped Newton method [50]. Here, we employ cross validation on the training dataset to find the initial value $\eta^{(0)}$ with Adagrad accordingly. Moreover, we update $V^{(t)}$ and $W^{(t)}$ as follows.

$$V^{(t)} = V^{(t-1)} - \eta^{(t)}(\Gamma(U^{(t-1)}, W^{(t-1)})^T \otimes I)^{-1} \nabla_{V^{(t-1)}} L,$$
$$W^{(t)} = W^{(t-1)} - \eta^{(t)}(\Gamma(U^{(t-1)}, V^{(t-1)})^T \otimes I)^{-1} \nabla_{W^{(t-1)}} L.$$

2SGD, which exploits a Hessian matrix, outperforms SGD because the optimal step size is equal to the inverse of the Hessian matrix when the surface is approximated as a quadratic plane. *STM-CP-2SGD* further utilizes the block diagonal components to approximate the Hessian matrix for acceleration. Notice that if the Hessian matrix is not invertible, a nonnegative diagonal matrix with negligible- valued elements can be added to the original matrix to produce a positive definite matrix [30].

Theoretical Results

In the following, we first derive the computational com - plexity of the above two algorithms. Afterward, we prove that *STM-CP-2SGD* always converges to the solution of CP decomposition. More specifically, let $|T|$ denote the number of nonzero elements in T .

Lemma 2. *The computational complexity of* STM-CP-SGD *is* $O((N + D + M)R^2 + N^2R + |T^{(t)}|R)$.

Proof. For V and W, the complexity of *STM-CP-SGD* is $O((D + M)R^2 + |T^{(t)}|R)$ for each update, where $O((D + M)R^2)$ is for computing $V\Gamma(U, W)$ and $W\Gamma(U, V)$, and $O(|T^{(t)}|R)$ is to find $T^{(t)}(U, , W)$ and $T^{(t)}(U, V,)$. For U, the complexity of *STM-CP-SGD* is $O(NR^2 + |T^{(t)}|R + N^2R)$ for each update since the time to find $L_a U$ is $O(N^2R)$. Therefore, the computational complexity of *STM-CP-SGD* is $O((N+D+M)R^2+N^2R+|T^{(t)}|R)$. The lemma follows. □

Lemma 3. *The computational complexity of* STM-CP-2SGD *is* $O((N + D + M)R^2 + N^2R + |T^{(t)}|R)$.

Proof. Compared with *STM-CP-SGD*, *STM-CP-2SGD* needs to additionally derive the inverse of the Hessian matrix. To update V, since $(\Gamma(U, W)^T \otimes I)^{-1} = (\Gamma(U, W))^{-1} \otimes I^{-1}$, we only have to additionally calculate the inverse of $\Gamma(U, W)$, and thus the computational complexity is $O(R^3)$, as well as for the update of W. On the other hand, for the update of U, efficiently calculating the inverse of the Kronecker sum, i.e., $(\Gamma(V, W)^T \oplus L_a) \in \mathbb{R}^{RN \times RN}$, has been studied in solving Sylvester equation in control system, and it can be approximated in $O(N^2 + R^2)$ [31]. Therefore, the computational complexity is still $O((N+D+M)R^2+N^2R+|T^{(t)}|R)$. The lemma follows. □

The worst-case computational complexity of *STM-CP-2SGD* is the same as *STM-CP-SGD*. Nevertheless, the com- putation of *STM-CP-2SGD* is slightly more complicated than that of *STM-CP-SGD* because *STM-CP-2SGD* needs to compute the inverse of the Hessian matrix. Neverthelss, as shown in the experimental results, the convergence rate of *STM-CP-2SGD* is much faster than that of *STM-CP-SGD*. Given the same start point, *STM-CP-2SGD* can adaptively control the learning step size to approach the optimal so- lution and thus requires much fewer iterations than that of *STM-CP-SGD* with a constant learning step size.

Assume that the Frobenius norm of T is bounded by a constant, i.e., $\|T\| \leq C$, we have the following theorem.

Theorem 1. *Given any initial solution or step size,* STM-CP-2SGD *does not diverge.*

Proof. STM-CP-2SGD updates U, V, W by finding the first and second derivatives. Consider the t-th update of $U^{(t)}$.
We first fix the values of $V^{(t-1)}$ and $W^{(t-1)}$ and set Eq. (11) as zero to find U^*, i.e.,

$$U(\Gamma(V^{(t-1)}, W^{(t-1)})+\lambda_2 I_R)+\lambda_1 L_a U = T_{(,V^{(t-1)}, W^{(t-1)})}. \quad (21)$$

Since Eq. (21) is a Sylvester-type equation, the optimal solution of U when fixing $V^{(t-1)}$ and $W^{(t-1)}$, denoted as U^*, can be derived as

$$T_{(,V^{(t-1)}, W^{(t-1)})}((\Gamma(V^{(t-1)}, W^{(t-1)})+\lambda_2 I_R)^T \oplus \lambda_1 L_a)^{-1},$$

ISBN 978-93-88808-61-3

which is the least-square solution by fixing $V^{(t-1)}$ and $W^{(t-1)}$. By rearranging Eq. (20), $U^{(t)}$ can be derived as

$$\begin{aligned}
&U^{(t-1)} - \eta^{(t)}((\Gamma(V^{(t-1)}, W^{(t-1)}) + \lambda_2 I_R)^T \oplus \lambda_2 L_a)^{-1} \\
&(-\nabla_U(\cdot, V^{(t-1)}, W^{(t-1)}) + U^{(t-1)}(\Gamma(V^{(t-1)}, W^{(t-1)}) + \lambda_2 I_R) \\
&+ \lambda_1 L_a U^{(t-1)}) \\
=&U^{(t-1)} + \eta^{(t)}\nabla_U(\cdot, V^{(t-1)}, W^{(t-1)})((\Gamma(V^{(t-1)}, W^{(t-1)}) \\
&+ \lambda_2 I_R)^T \oplus \lambda_2 L_a)^{-1} - \eta^{(t)} U^{(t-1)} \\
=&(1-\eta^{(t)}) U^{(t-1)} + \eta^{(t)} U^*.
\end{aligned}$$

Therefore, $U^{(t)}$ is a linear combination of $U^{(t-1)}$ and U^*. When $V^{(t-1)}$ and $W^{(t-1)}$ are fixed as constants, according to [53], U^* is a solution of the least-square problem, and thus

$$L(T, U^*, V^{(t-1)}, W^{(t-1)}) \leq L(T, U, V^{(t-1)}, W^{(t-1)}), \forall U.$$

Let U in the right-hand side be 0, and we have

$$L(T, U^*, V^{(t)}, W^{(t)}) \leq \frac{\eta^{(t)}}{2} \|T\|^2. \quad (0)$$

Therefore, since the Frobenius norm of T is bounded, the initial solution of U is also bounded, and so is $U^{(t)}$. The updates of $V^{(t)}$ and $W^{(t)}$ are similar. The theorem follows. □

Notice that this theorem does not guarantee the convergence[9] of the algorithm since the solutions may oscillate when updating U, V, and W sequentially. Nevertheless, the above property is useful in practice since the solution can be always bounded and does not result in overflow [54].

5 Experimental Results

In this section, we evaluate SNMDD with real datasets. A user study with 3126 people is conducted to evaluate the accuracy of SNMDD. Moreover, a feature study is performed. Finally, we apply SNMDD on large-scale datasets and analyze the detected SNMD types.

Data Preparation and Evaluation Plan

In the following, we detail the preparation of the datasets used in our evaluation.

User Study

We recruit 3126 OSN users around the world via Amazon Mechanical Turk (MTurk) to obtain data for training and testing the classifiers in SNMDD. The participants include 1790 males and 1336 females. Their professions are very diverse, affiliating with universities, government offices, technology companies, art centers, banks, and businesses. Each user is first invited to fill out the standard SNMD questionnaires [1], [18].[10] Then, a group of professional psychiatrists participating in this project assess and manually label the users as *potential SNMD cases* (and their types of SNMDs) or *normal users*.[11] There are 389 users labeled as *SNMD*, including 246 Cyber-Relationship (CR) Addiction, 267 Information Overload (IO), and 73 Net Compulsion (NC).[12] The result obtained by the psychiatrists serves as the ground truth for our evaluation. We also crawl the Facebook (denoted as FB_US) and Instagram (denoted as IG_US) data of the participants in the user study for training and testing of our machine learning models (based on features detailed in Section 3.1). All the data are collected with the Facebook and Instagram APIs as listed in Table 1.

TABLE 1
Details of the datasets

Dataset	Description
FB_US	User profile, the friends of each user, the news feeds created by users with metadata (who likes, who comments, stickers, and geotag), the news feeds users like orcomment (stickers also), events (join/decline), joined groups with events, and game posts created by game apps
IG_US	User profile, the followers/followees of each user, the media created by users with metadata (who likes, who comments, and geotag), and the contents users like or comment
FB_L	Anonymized user ID that performs the action, anonymized user ID that receives the action, and timestamp of action creation
IG_L	Anonymized media ID, anonymized ID of the user who created the media, timestamp of media creation, set of tags assigned to the media, number of likes and number of comments received

In the experiment, we first evaluate the effectiveness of the proposed features, including all features (All), social interaction features (Social), personal profile features (Personal), with a baseline feature Duration, i.e., the total time spent online, using TSVM [13] for semi-supervised learning in the user study. The combinations of different features, i.e., Duration with Social (D-S), Duration with Personal (D-P), Social with Personal (S-P), are also presented. We also collect two large-scale datasets, including Facebook (denoted as FB_L) with 63K nodes, 1.5M edges, and 0.84M wall posts [19], and Instagram (denoted as IG_L) with 2K users, 9M tags, 1200M likes, and 41M comments [20]. Note that some proposed features cannot be extracted from certain large-scale datasets, e.g., game posts and stickers are not available in IG_L, which is handled by using the imputation technique [21]. The details of the data crawled from each social media are listed in Table 1.

With labeled (IG_US and FB_US) and unlabeled data (IG_L and FB_L) described above, we perform a 5-fold cross validation, i.e., take 4 folds for training and 1 fold for testing, to evaluate the performance of proposed features using semi-supervised TSVM. A number of supervised approaches, including J48 Decision Tree Learning [22], SVM [23], Logistic Regression, and DTSVM [24] which do not use unlabeled data, are also compared to justify our choice of using TSVM in SNMDD. Next, we compare the proposed *SNMD-based Tensor Model (STM)*, implemented by different algorithms, i.e., *STM-Tucker-SGD*, *STM-CP-SGD*, and *STM-CP-2SGD*, with two baseline algorithms. The first baseline algorithm concatenates the features from different networks together (denoted as *CF*), while the second baseline algorithm employs the existing Tucker model (denoted as *Tucker*) that does not incorporate prior knowledge regarding the characteristics of SNMD cases (observed from our analysis).

9. To ensure the convergence of 2SGD, more assumptions, e.g., $(\alpha, \gamma, s, \delta)$-strict saddle with stochastic gradient oracle with the radius at most Q, are required to prove the range to optimal solutions [51], [52], but the above condition may not always exist in practical situations.

10. The IRB number of this project is AS-IRB-HS 15003 v.1.

11. They are from California School of Professional Psychology, Taipei City Hospital, Nat'l Taipei Univ., psychiatric clinics, etc.

12. Note that a person may have multiple types of SNMDs simultaneously.

Finally, the effectiveness of each feature is carefully analyzed in Section 5.5.

Large-Scale Experiments

To discover new insights, we apply our semi-supervised SNMDD on IG_L and FB_L to classify their users and then analyze the detected cases of different SNMD types. Notice that the goal of this analysis is exploratory-oriented as we do not have the ground truth for the large datasets. We examine whether friends of SNMD cases tend to be potential SNMD cases as well. Also, we apply community detection on FB L and IG_L to derive the relationships between different types of SNMD users in their communities. Finally, the average hop distance between the SNMD users of the same type is reported.

Evaluation of the Proposed Features

In the following, we first evaluate the performance of the proposed features using TSVM. We adopt Accu- racy (Acc.) and Area Under Curve (AUC) for evalua- tion of SNMDD. Moreover, Microaveraged-F1 (Micro-F1) and Macroaveraged-F1 (Macro-F1) are also compared for multiple-label classification. Table 2 summarizes the average results and standard deviations, where the examined feature sets are denoted by self-explained labels.

The results on the IG_US and FB_US datasets in the user study show that Durationleads to the worst performance, i.e., the results of accuracy are 34% and 36%, and the AUC are 0.362 and 0.379, respectively. Notice that the AUC function can flip the results if the calculated AUC is less than 0.5, i.e., 1-AUC. Here, we do not filp the results to show that Durationis in fact a negative predictor in our case because Duration cannot differentiate heavy users with addictive users. Using all (All) or parts (Social or Personal) of the proposed features outperforms Duration significantly (see Table 2). All achieves the best performance (80% and 84% accuracy on the IG_US and FB_US datasets, respectively) because SNMDD is able to capture the various features ex- tracted from data logs to effectively detect SNMD cases. The performance of Personal and Social are comparable, and the integrated feature set All outperforms Personal and Social by at least 15% and 16% on IG_US and FB_US in terms of accuracy. Since the F1 measure ignores true negatives, its magnitude is mostly determined by the number of true positives, i.e., large classes dominate small classes in microaveraging. As shown in Table 2, Micro-F1 of Duration, Social, and Personal are larger than Macro- F1 using both IG_US and FB_US datasets, indicating that using parts of features performs better on IO and CR (large classes) than NC. In contrast, the performance of SNMDD is almost the same in Micro-F1 and Macro-F1, which indicates its robustness. The results from FB_US are better than those from IG_US because IG_US is sparser, e.g., there are no event and game posts on Instagram. After comparing the results from SNMDD with the ground truth obtained via user study, we observe that some false-positive users are detected as NC, probably because people with NC are more likely to hide their real usage time, e.g., the game logs of some people with NC are hidden. As a result, a few normal users may be incorrectly detected as NC. However, SNMDD generally performs very well for NC due to some effective features. For example, users of NC are usually less parasocial since they are less frequent to interact with friends. Moreover, since the NC users' friends with game benefits usually do not know the NC users' other friends (e.g., colleagues), their clustering coefficients are lower than the normal users. Finally, the performance of Social and Personal with Duration features (D-S and D-P) are almost the same since SVM finds the best hyperplane to classify the training data and may not take the dimensions that downgrade the results. The results also manifests that the proposed features are robust with SVM. The p-value tests of different features with All indicate that All is significantly better than Duration, Social, Personal, D-S, D-P with p values that are much smaller than 0.05, while the performance is close to S-P.

Evaluation of Classification Techniques and STM

In the following, given all the proposed features, we first evaluate TSVM in comparison with some representative supervised learning approaches in SNMDD. As shown in Table 3, the accuracy of semi-supervised TSVM (84.3%) outperforms all the supervised algorithms, including A_2- regularized logistic regression (78.6%) and A_2-regularized A_2-loss SVM (79.2%), since TSVM effectively uses unlabeled data to address the issues of overfitting and data sparsity. The accuracy and AUCs of the single-source supervised learning methods are similar, indicating that the proposed features provide robust information that is not sensitive to the choice of learning algorithms.

Next, we compare the proposed multi-source *STM- Tucker-SGD*, *STM-CP-SGD*, and *STM-CP-2SGD* with two baselines, i.e., *CF* and *Tucker*, to integrate the features ex- tracted from IG_US and FB__US datasets with TSVM. Table 3 points out that the accuracy and AUC of *STM-CP-2SGD* are 90.4% and 0.938, respectively. *STM-CP-2SGD* with the decomposed latent factor matrix U can effectively recover important missing features and provide extra latent infor- mation to better characterize the users. In contrast, *CF*, which simply concatenates the features from FB_US and IG_US, suffers from the worst accuracy and AUC and is even beated by single-source A_2-regularized A_2-loss SVM. This is because *CF* loses correlations among the features and thereby tends to introduce noises. On the other hand, *STM-CP-2SGD* outperforms the other approaches because it incorporates important characteristics of SNMD and thereby derives more precise and accurate latent features, while the accuracy and AUC of *STM-Tucker-SGD* and *STM-CP-SGD* are almost the same.

Evaluation of the Proposed Tensor Decomposition Acceleration

In the following, the default dimensionality of U, V and W, threshold *s*, and the maximum number of iterations are set as 10, 0.001, and 50, respectively. We first compare the loss function through each iteration in Figure 2(a). Note that *STM-CP-2SGD* converges very quickly (always terminates before the 5-th iteration). Between *STM-Tucker- SGD* and *STM-CP-SGD*, the loss of *STM-Tucker-SGD* is slightly smaller since the core tensor of *STM-Tucker-SGD*

ISBN 978-93-88808-61-3

TABLE 2
Different combinations of feature categories for performance evaluations on the IG US and FB_US datasets.

Instagram Measure	Duration	Social	Personal	D-S	D-P	S-P	All
Acc.	0.34±0.02	0.67±0.01	0.69±0.03	0.63±0.02	0.68±0.02	0.80±0.01	0.80±0.01
AUC	0.36±0.02	0.71±0.02	0.74±0.01	0.69±0.02	0.73±0.02	0.81±0.01	0.81±0.01
Micro-F1	0.33±0.01	0.76±0.01	0.78±0.04	0.74±0.01	0.77±0.03	0.85±0.01	0.85±0.01
Macro-F1	0.33±0.01	0.71±0.01	0.73±0.02	0.69±0.02	0.72±0.02	0.85±0.01	0.85±0.01
p value on AUC	$3.80 \cdot 10^{-8}$	$6.05 \cdot 10^{-5}$	$1.18 \cdot 10^{-5}$	$1.64 \cdot 10^{-6}$	$3.06 \cdot 10^{-6}$	0.76	-
Facebook							
Acc.	0.36±0.01	0.72±0.03	0.73±0.02	0.70±0.03	0.73±0.01	0.84±0.02	0.84±0.02
AUC	0.37±0.01	0.75±0.02	0.77±0.02	0.74±0.01	0.76±0.02	0.86±0.02	0.85±0.01
Micro-F1	0.44±0.04	0.80±0.02	0.81±0.01	0.79±0.02	0.80±0.01	0.90±0.01	0.90±0.01
Macro-F1	0.35±0.02	0.76±0.01	0.77±0.03	0.74±0.01	0.76±0.02	0.91±0.01	0.91±0.01
p value on AUC	$7.01 \cdot 10^{-9}$	$2.35 \cdot 10^{-5}$	$3.06 \cdot 10^{-4}$	$3.90 \cdot 10^{-6}$	$1.46 \cdot 10^{-4}$	0.064	-

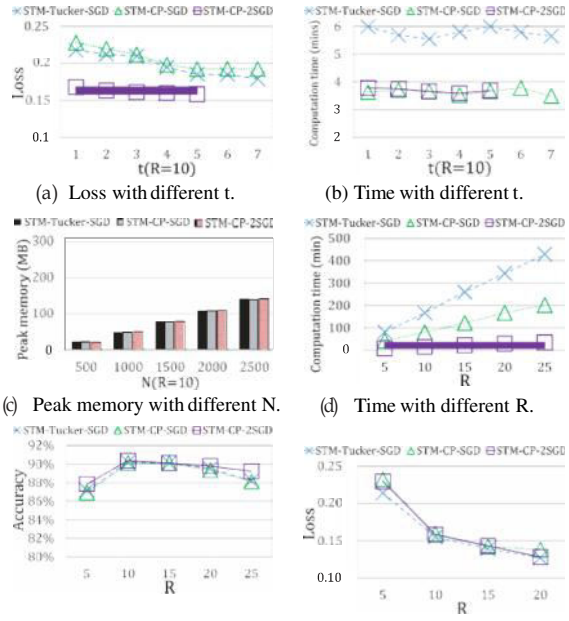
TABLE 3
Comparisons of SNMDD with different classification techniques.

Technique	Acc.	AUC
Single-source (FB)		
J48 Decision Tree Learning	75.4%	0.763
A_1-regularized A_2-loss SVM	79.1%	0.790
A_2-regularized A_2-loss SVM	79.2%	0.791
A_1-regularized logistic regression	78.5%	0.788
A_2-regularized logistic regression	78.6%	0.789
DTSVM	78.5%	0.782
TSVM	**84.2%**	**0.851**
Multi-source (FB+IG)		
CF	76.4%	0.775
Tucker	87.9%	0.892
STM-Tucker-SGD	**90.2%**	**0.933**
STM-CP-SGD	**90.1%**	**0.930**
STM-CP-2SGD	**90.4%**	**0.938**

allows more freedom to fit data. On the other hand, *STM- CP-2SGD* outperforms both *STM-Tucker-SGD* and *STM-CP- SGD* in terms of the convergence rate, as well as the loss function. Figure 2(b) compares the running time through each iteration. The results manifest that the running time of *STM-CP-SGD* is the fastest for each iteration, with *STM-CP-*

2SGD as the close second (which terminates first). They both significantly outperforms *STM-Tucker*. The overall running time of *STM-CP-2SGD* is the smallest since it requires much fewer iterations. Figure 2(c) shows the peak memory usage of different tensor decomposition methods. The memory usage of *STM-CP-2SGD* is slightly greater than that of the others. Notice that the feature tensor and adjacency matrix are sparse, and therefore we use sparse tensor representa- tion for each decomposition to reduce the memory usage.

Moreover, Figures 2(d)-(f) compare the performance of *STM-Tucker-SGD*, *STM-CP-SGD*, and *STM-CP-2SGD* in terms of running time, accuracy, and loss function with different *R*, respectively. As shown in Figures 2(d), *STM- CP-2SGD* significantly outperforms the other two in terms of running time for different *R*, i.e., the running time is at most 9.7% and 18.9% of *STM-Tucker-SGD* and *STM-CP-SGD*, re- spectively, while the accuracy for detecting SNMD is almost the same for different proposed methods as shown in Figure 2(e), which shows the power of acceleration of the proposed *STM-CP-2SGD* without sacrificing accuracy. Moreover, the accuracy of different methods does not increase as *R* grows since the latent features may overfit the training data and thus do not perform well on testing data. Figure 2(f) further shows the loss function with different *R*. As *R* increases, the loss functions for different methods all decrease. For



(a) Loss with different t. (b) Time with different t.

(c) Peak memory with different N. (d) Time with different R.

(e) Accuracy with different R. (f) Loss with different R.

Fig. 2. Comparisons of different datasets.

Figure 2(e) and Figure 2(f), although the loss function of *STM-Tucker-SGD* slightly outperforms *STM-CP-2SGD*, the accuracy of *STM-Tucker-SGD*, *STM-CP-SGD*, and *STM-CP- 2SGD* are similar. In summary, *STM-CP-2SGD* is the most efficient one without compromising efficiency and accuracy.

Feature Study

To observe the differences among the three types of SN- MDs, Table 4 lists the top-5 discriminative features using information gain and the corresponding accuracy on the FB_US dataset by TSVM, where CC, BI, BL, and SD re- spectively denote the clustering coefficient, burst intensity, burst length, and standard deviation. It is worth noting that the number of selfies, an indicator of self-disclosure, is not useful for detecting CR and IO, but it is effective for NC. This is because NC users are usually less socially active, comparing to CR and IO users. Moreover, the online/offline interaction ratio of NC is much higher than the ratios of the other two types, probably because NC users usually show less willingness to join offline activities. In contrast, CR

ISBN 978-93-88808-61-3

and IO cases prefer to use social media, instead of playing games alone. Moreover, people with compulsive personality are more introverted. In contrast, people with CR usually create virtual bonds to develop pathological relationships for compensation of their (missing) offline relationships. The Shannon index, an indicator of the social diversity, is also

useful in detecting NC since the friends of NC are in similar backgrounds, and thus the Shannon index is lower than that of normal users. Moreover, the social comparison score

is important for detecting NC cases. This is because when the users with malicious envy see the positive newsfeeds from the friends with similar backgrounds, they may eager

to pursue the sense of success, which is much easier to be achieved in online games.

The parasociality, effective for detecting all SNMD types, is especially useful for detecting CR cases. For example, in our user study, we find user A, 21-year-old male, frequently posting news feeds, such as "I'm so bored :((((...Ahhh- hhh!!", and his cross-dressing photos on his Facebook time-

line, more than 3 times a week, which usually get fewer than 5 likes. At the same time, he "likes" a large number of posts from others. SNMDD classifies him as a potential CR case and his questionnaire reveals that he constantly blocks out

disturbing thoughts about life and finds himself anticipating when he goes online again.

Burst intensity and length are quite useful for detecting IO cases. For example, user B, 36-year-old male, is detected as IO since the behavior of clicking "likes" fits the pattern of bursts, i.e., the median of his burst intensity is high, equal to 31. His answers to the standard questionnaire reveal that he loses sleep due to late-night access on Facebook to check others' news feeds. Through interview, user B explains that he cannot stop checking for new posts and e-mails even when all his news feeds and emails are read. Some of his friends reply him: "are you a robot? no sleep needed?!?!!", indicating that user B is indulged in finding social news. Moreover, social roles are important in detecting IO since the users with IO usually share or like the information from different communities ant thus are inclined to be detected as structural holes.

Next, we analyze the importance of different features

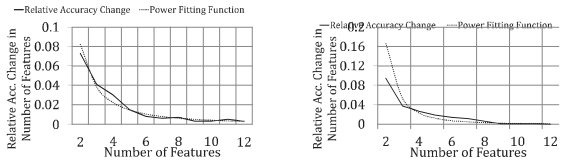
to our classifiers. χ^2-test is exploited to measure the impor- tance of each feature via SelectKBest of Scikit-Learn. The top 5 important features overall are 1) median of the intensity of bursts, 2) parasociality, 3) online/offline interaction ratio,

4) number of used stickers, and 5) standard deviation of the length of bursts. It is worth noting that TSVM using only these 5 features in SNMDD achieves an accuracy of 80.7% for FB_ US, close to that of using all features (All). In other words, integrating important social and personal features provides good results because effective personal features, e.g., the temporal behavior features, can be used to differentiate the users suffering from withdraw orrelapse symptoms and heavy users, while social features capture the interactions among users to differentiate different SNMDs.

Figs. 3(a) and 3(b) show the improvement made by adding different features in TSVM on the FB_US dataset and the proposed *STM* on multi-source data (i.e., FB_US and IG_US). The feature selection of TSVM is based on the information gain (the top-5 features mentioned earlier), while the tensor approach automatically extracts important

TABLE 4
Top features and Acc. on the FB US dataset.

CR	NC	IO	
Parasociality	Game posts	Median of BI	Median of BI
Median of BI	Online/offline ratio	Online/offline ratio	Online/offline ratio
Eigenvector centrality	Parasociality	SD of BL	Parasociality
Online/offline ratio	Shannon index	Sticker number	Sticker number
Sticker number	Social comparison score	Social roles	SD of BL
Acc.: 80.5%	Acc.: 77.6%	Acc.: 82.9%	Acc.: 80.7%



(a) Relative improvement w.r.t. the number of features.

(b) Relative improvement w.r.t. the number of latent features (*STM*).

Fig. 3. Relative accuracy change with respect to number of features.

latent features. We observe a diminishing return property on both figures, where the improvement becomes marginal as more features are included. Fig. 3(a) shows a power fit

function ($p(x) = 0.3091x^{-1.92}$) of the curve with $R^2 = 0.9534$. The exponent 1.92 denotes that the improvement by adding n-th feature is $n^{-1.92}$ times smaller than that by adding the first feature. On the other hand, the results

of the tensor-based approach in Fig. 3(b) show that the accuracy increment for adding a single feature drops faster

($p(x) = 1.11x^{-2.01}$) since the proposed *STM* can extract much more important and concise features.

Analysis of SNMD Types in Large Datasets

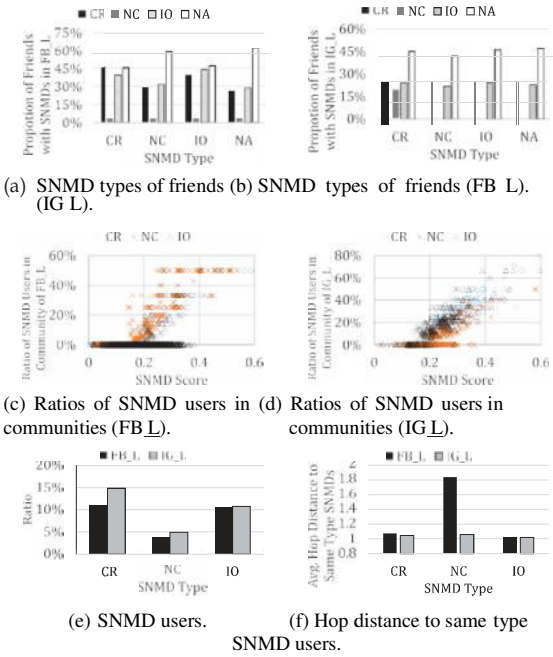
In this analysis, we first apply the proposed SNMDD frame- work (with TSVM) on some large-scale OSN datasets, i.e., FB_L and IG_L, to classify their users. In Figs. 4(a) and 4(b), we analyze the detected SNMD cases among the friends of an SNMD user. In Fig. 4(a), the leftmost bar indicates that in FB_L, among all CR users, about 45% of their friends are also CR users, which is greater than the percentage of other SNMD types. On the other hand, the 8th bar from the left in Fig. 4(a) indicates that in FB_L, about 59% of NC users' friends are NA (non-SNMD users). Figs. 4(a) and 4(b) show that, in FB_L and IG_L, CR and IO users have similar friend types. This is because CR and IO cases, by their nature, are similar, i.e., they are both seeking social satisfaction (e.g., relationships and information) from the OSNs. Moreover, among different SNMD cases, CR and IO users are likely to be friends with other CR and IO users. For CR users, this phenomenon has been described as "loneliness propagates" [15].

Furthermore, Infomap community detection [39] is per- formed on FB_L and IG_L to derive the relationships be- tween different types of SNMD users in their communities. Figs. 4(c) and 4(d) analyze the community structures of

ISBN 978-93-88808-61-3

SNMD users with different SNMD scores, where each point represents the characteristic of a community. Specifically, each community in the dataset is represented by three differ- ent types of points, i.e., CR, NC, and IO. For example, each CR point is represented as *(score, ratio)*, where *score* is the average CR score in that community, and *ratio* indicates the proportion of CR users in the community. It is similar for each IO/NC point. As Figs. 4(c) and 4(d) show, for each SNMD type, when the average SNMD score is higher, it is likely to have more SNMD users in the community. Moreover, there are many communities with large IO scores in IG_L that have IO ratios close to 1. This implies that the users with large IO scores in IG_L are more inclined to form homogeneous groups. At the first glance, one may feel that NC users frequently appear in many communities, and there seems to be a large number of NC users, especially in FB_L (i.e., Fig. 4(c)). However, after carefully examining these communities, we find that those communities (with large ratios of NC users) are usually very small (usually with the size around 5) because NC users are less-active. On the other hand, in IG_L, when SNMD scores are larger, the ratios of IO users in communities are also larger. This is because IO users can view, like, or follow others in Instagram more easily (not necessary to be friends first).

Fig. 4(e) compares the ratios of different types of SNMD users identified in FB_L and IG_L. There are more CR users in IG_L probably because CR users seek social supports online to compensate the loneliness in real life. We argue that the Instagram platform makes it easy to freely create social relationships with strangers. In contrast, it is not that easy to create new social relationships on Facebook since the friend requests need to be approved. Finally, Fig. 4(f) compares the average number of hops from each SNMD user to the nearest user with the same type of SNMDs. The leftmost bar shows that the average hop distance from each CR user to the closest CR user is 1.07 hop, indicating that CR and IO users are close to other same-type users, i.e., average hop distances are within 1.15, where Figs. 4(a) and 4(b) also report similar results.



(a) SNMD types of friends (FB_L). (b) SNMD types of friends (IG_L).

(c) Ratios of SNMD users in communities (FB_L). (d) Ratios of SNMD users in communities (IG_L).

(e) SNMD users. (f) Hop distance to same type SNMD users.

Fig. 4. Comparisons of different datasets.

6 Conclusion

In this paper, we make an attempt to automatically identify potential online users with SNMDs. We propose an SNMDD framework that explores various features from data logs of OSNs and a new tensor technique for deriving latent fea- tures from multiple OSNs for SNMD detection. This work represents a collaborative effort between computer scien- tists and mental healthcare researchers to address emerging issues in SNMDs. As for the next step, we plan to study the features extracted from multimedia contents by techniques on NLP and computer vision. We also plan to further explore new issues from the perspective of a social network service provider, e.g., Facebook or Instagram, to improve the well-beings of OSN users without compromising the user engagement.

References

[1] K. Young, M. Pistner, J. O'Mara, and J. Buchanan. Cyber-disorders: The mental health concern for the new millennium. *Cyberpsychol. Behav.*, 1999.

[2] J. Block. Issues of DSM-V: internet addiction. *American Journal of Psychiatry*, 2008.

[3] K. Young. Internet addiction: the emergence of a new clinical disorder, *Cyberpsychol. Behav.*, 1998.

[4] I.-H. Lin, C.-H. Ko, Y.-P. Chang, T.-L. Liu, P.-W. Wang, H.-C. Lin, M.-F. Huang, Y.-C. Yeh, W.-J. Chou, and C.-F. Yen. The associa- tion between suicidality and Internet addiction and activities in Taiwanese adolescents. *Compr. Psychiat.*, 2014.

[5] Y. Baek, Y. Bae, and H. Jang. Social and parasocial relationships on social network sites and their differential relationships with users' psychological well-being. *Cyberpsychol. Behav. Soc. Netw.*, 2013.

[6] D. La Barbera, F. La Paglia, and R. Valsavoia. Social network and addiction. *Cyberpsychol. Behav.*, 2009.

[7] K. Chak and L. Leung. Shyness and locus of control as predictors of internet addiction and internet use. *Cyberpsychol. Behav.*, 2004.

[8] K. Caballero and R. Akella. Dynamically modeling patients health state from electronic medical records: a time series approach. *KDD*, 2016.

[9] L. Zhao and J. Ye and F. Chen and C.-T. Lu and N. Ramakr- ishnan. Hierarchical Incomplete multi-source feature learning for Spatiotemporal Event Forecasting. *KDD*, 2016.

[10] E. Baumer, P. Adams, V. Khovanskaya, T. Liao, M. Smith, V. Sosik, and K. Williams. Limiting, leaving, and (re)lapsing: an exploration of Facebook non-use practices and experiences. *CHI*, 2013.

[11] R. Jain and N. Abouzakhar. A comparative study of hidden markov model and support vector machine in anomaly intrusion detection. *JITST*, 2013.

[12] C. Tan, L. Lee, J. Tang, L. Jiang, M. Zhou, and P. Li. User-level sentiment analysis incorporating social networks. *KDD*, 2011.

[13] R. Collobert, F. Sinz, J. Weston, and L. Bottou. Large scale trans- ductive svms. *JMLR*, 2006.

[14] L. Leung. Net-generation attributes and seductive properties of the internet as predictors of online activities and internet addic- tion. *Cyberpsychol. Behav. Soc. Netw.*, 2004.

[15] J. Cacioppo, J. Fowler, and N. Christakis. Alone in the crowd: the structure and spread of loneliness in a large social network. *J. Pers. Soc. Psychol.*, 2009.

[16] J. Kleinberg. Bursty and hierarchical structure in streams. *KDD*, 2002.

[17] K.-L. Liu, W.-J. Li, and M. Guo. Emoticon smoothed language models for twitter sentiment analysis. *AAAI*, 2012.

[18] C. Andreassen, T. Torsheim, G. Brunborg, and S. Pallesen. Devel- opment of a Facebook addiction scale. *Psychol. Rep.*, 2012.

[19] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi. On the evolution of user interaction in Facebook. *WOSN*, 2009.

[20] E. Ferrara, R. Interdonato, and A. Tagarelli. Online popularity and topical interests through the lens of Instagram. *HT*, 2014.

[21] M. Saar-Tsechansky and F. Provost. Handling missing values when applying classification models. *JMLR*, 2007.
[22] I. Witten and E. Frank. Data mining: practical machine learn- ing tools and techniques with Java implementations. Morgan- Kaufmann, San Francisco, 2000.
[23] C.-C. Chang and C.-J. Lin. LIBSVM: a library for support vector machines, 2001.
[24] F. Chang, C.-Y. Guo, X.-R. Lin, and C.-J. Lu. Tree decomposition for large-scale SVM problems. *JLMR*, 2010.
[25] P. Comon, X. Luciani, and A. L. D. Almeida. Tensor decomposi- tions, alternating least squares and other tales. *Journal of Chemo- metrics*, 2009.
[26] E. Acar, D. M. Dunlavy, and T. G. Kolda. A scalable optimization approach for fitting canonical tensor decompositions. *Journal of Chemometrics*, 2011.
[27] L. He, C.-T. Lu, J. Ma, J. Cao, L. Shen, and P. S. Yu. Joint community and structural hole spanner detection via harmonic modularity. *KDD*, 2016.
[28] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummad. Measur- ing user influence on twitter: The million follower fallacy. *ICWSM*, 2010.
[29] K. Hayashi, T. Maehara, M. Toyoda, and K. Kawarabayashi. Real- time top-r topic detection on twitter with topic hijack filtering. *KDD*, 2015.
[30] J. Gill and G. King. What to do when your Hessian is not invert- ible: Alternatives to model respecification in nonlinear estimation. *Sociological Methods and Research*, 2004.
[31] B. Kågström and P. Poromaa. Distributed and shared memory block algorithms for the triangular Sylvester equation with sep^{-1} estimators. *SIAM Journal on Matrix Analysis and Applications*, 1992.
[32] L. R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 1966.
[33] R.A. Harshman. Foundations of the PARAFAC procedure: Models and conditions for an explanatory multimodal factor analysis. *UCLA Working Papers in Phonetics*, 1970.
[34] S. Graham, A. Munniksma, J. Juvonen. Psychosocial benefits of cross-ethnic friendships in urban middle schools. *Child Develop- ment*, 2013.
[35] S. S. Levine, E. P. Apfelbaum, M. Bernard, V. L. Bartelt, E. J. Zajac, and D. Stark. Ethnic diversity deflates price bubbles. *National Academy of Sciences*, 2014.
[36] H. Appel, J. Crusius, and Alexander L. Gerla. Social comparison, envy, and depression on facebook: a study looking at the effects of high comparison standards on depressed individuals. *Journal of Social and Clinical Psychology*, 2015.
[37] A. Bordes, L. Bottou, and P. Gallinari. SGD-QN: careful quasi- newton stochastic gradient descent. *Journal of Machine Learning Research*, 2009.
[38] J. B. White, E. J. Langer, L. Yariv, and J. C. Welch IV. Frequent social comparisons and destructive emotions and behaviors: the dark side of social comparisons. *Journal of Adult Development*, 2006.
[39] M. Rosvall and C. Bergstrom. Maps of random walks on complex networks reveal community structure. *Natl. Acad. Sci.*, 2008.
[40] D. L. King, P. H. Delfabbro, D. Kaptsis, and T. Zwaans. Adolescent simulated gambling via digital and social media: an emerging problem. *Computers in Human Behavior*, 2014.
[41] D. Li, X. Li, L. Zhao, Y. Zhou, W. Sun, and Y. Wang. Linking multiple risk exposure profiles with adolescent Internet addiction: insights from the person-centered approach. *Computers in Human Behavior*, 2017.
[42] K. Kim, H. Lee, J. P. Hong, M. J. Cho, M. Fava, D. Mischoulon, D. J. Kim, and H. J. Jeon. Poor sleep quality and suicide attemptamong adults with internet addiction: a nationwide community sample of Korea. *PLOS ONE*, 2017.
[43] C.-H Chang, E. Saravia, and Y.-S. Chen. Subsonscious crowd- sourcing: a feasible data collection mechanism for mental disorder detection on social media, *ASONAM*, 2016.
[44] B. Saha, T. Nguyen, D. Phung, and S. Venkatesh. A framework for classifying online mental health-related communities with an interest in depression. *IEEE Journal of Biomedical and Health Informatics*, 2016.
[45] M. Choudhury, M. Gamon, S. Counts, and E. Horvitz. Predicting depression via social media. *ICWSM*, 2013.
[46] T. Kolda and B. Bader. Tensor decompositions and applications. *SIAM review*, 2009.
[47] H.-H. Shuai, C.-Y. Shen, D.-N. Yang, Y.-F. Lan, W.-C. Lee, P. S. Yu, and M.-S. Chen. Mining online social data for detecting social network mental disorders. *WWW*, 2016.
[48] A. Anandkumar, D. Hsu, M. Janzamin, and S. Kakade. When are overcomplete topic models identifiable? uniqueness of tensor tucker decompositions with structured sparsity. *JLMR*, 2015.
[49] L. Bottou. Stochastic gradient descent tricks. *Neural Networks: Tricks of the Trade*, 2012.
[50] J. M. Ortega and W. C. Rheinboldt. Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, NY, 1970.
[51] Y. Nesterov and B. T. Polyak. Cubic regularization of Newton method and its global performance. *Mathematical Programming*, 2006.
[52] R. Ge, F. Huang, C. Jin, and Y. Yuan. Escaping from saddle pointsonline stochastic gradient for tensor decomposition. *Confer- ence of Learning Theory (COLT)*, 2015.
[53] T. Maehara, K. Hayashi, and K. Kawarabayashi. Expected tensor decomposition with stochastic gradient descent. *AAAI*, 2016.
[54] V. Silva and L.-H. Lim. Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM Journal on Matrix Analysis and Applications*, 2006.

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

Nearest Keyword Set Search in Multi-Dimensional Datasets

Srinivas Reddy,
M.Tsech
Scholar, CSE
Department,
Malla Reddy College of Engineering

Abstract—Keyword-based search in text-rich multi-dimensional datasets facilitates many novel applications and tools. In this paper, we consider objects that are tagged with keywords and are embedded in a vector space. For these datasets, we study queries that ask for the tightest groups of points satisfying a given set of keywords. We propose a novel method called ProMiSH (Projection and Multi Scale Hashing) that uses random projection and hash-based index structures, and achieves high scalability and speedup. We present an exact and an approximate version of the algorithm. Our experimental results on real and synthetic datasets show that ProMiSH has up to 60 times of speedup over state-of-the-art tree-based techniques.

Index Terms—Querying, multi-dimensional data, indexing, hashing

1 INTRODUCTION

OBJECTS (e.g., images, chemical compounds, documents, or experts in collaborative networks) are often charac- terized by a collection of relevant features, and are com- monly represented as points in a multi- dimensional feature space. For example, images are represented using color fea- ture vectors, and usually have descriptive text information (e.g., tags or keywords) associated with them. In this paper, we consider multi-dimensional datasets where each data point has a set of keywords. The presence of keywords in feature space allows for the development of new tools to query and explore these multi-dimensional datasets.

In this paper, we study *nearest keyword set* (referred to as NKS) queries on text-rich multi-dimensional datasets. An NKS query is a set of user-provided keywords, and the result of the query may include *k* sets of data points each of which contains all the query keywords and forms one of the top-*k* tightest cluster in the multi- dimensional space. Fig. 1 illustrates an NKS query over a set of two-dimensional data points. Each point is tagged with a set of keywords. For a query *Q* ¼ f*a; b; c*g, the set of points f7*;* 8*;* 9g contains all the query keywords f*a; b; c*g and forms the tightest cluster com- pared with any other set of points covering all the query keywords.

Therefore, the set f7*;* 8*;* 9g is the top-1 result for the query *Q*.

NKS queries are useful for many applications, such as photo-sharing in social networks, graph pattern search, geo- location search in GIS systems[1] [1], [2], and so on. The fol- lowing are a few examples.

1) Consider a photo-sharing social network (e.g.,Facebook), where photos are tagged with people names and locations. These photos can be embed- ded in a high-dimensional feature space of texture, color, or shape [3], [4]. Here an NKS query can find a group of similar photos which contains a set of people.
2) NKS queries are useful for graph pattern search, where labeled graphs are embedded in a high dimensional space (e.g., through Lipschitz embed- ding [5]) for scalability. In this case, a search for a subgraph with a set of specified labels can be answered by an NKS query in the embedded space [6].
3) NKS queries can also reveal geographic patterns. GIS can characterize a region by a high-dimensional set of attributes, such as pressure, humidity, and soil types. Meanwhile, these regions can also be tagged with information such as diseases. An epidemiolo- gist can formulate NKS queries to discover patterns by finding a set of similar regions with all the dis- eases of her interest.

We formally define NKS queries as follows.

Nearest keyword set. Let Dc R^d be a *d*-dimensional data-set with *N* points. For any *o* 2 D, it is tagged with a set of keywords s*ðo*Þ=fv_1*; :::;* v_tg ≤ V, where V is a dictionary of *U* unique keywords. For any o_i*;* o_j 2 D, the distance between o_i and o_j is measured by their L_2-norm (i.e., euclidean distance) as *dist*ðo_i*;* o_jÞ ¼ ko_i — o_jk$_2$. Given a set

ISBN 978-93-88808-61-3

of data points A c D, rðA Þ is the diameter A and is defined by the maximum distance between any two points in A,

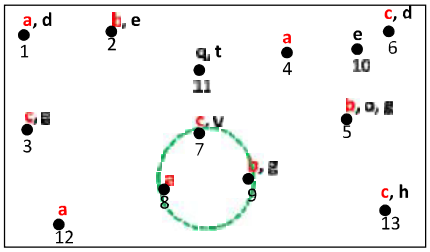


Fig. 1. An example of an NKS query on a keyword tagged multi-dimen- sional dataset. The top-1 result for query $\{a, b, c\}$ is the set of points $\{7, 8, 9\}$.

A smaller rðA Þ implies the points in A are more similar to each other.

Given an NKS query with q keywords $Q = \{v_1, \ldots, v_q\}$, $A \subseteq D$ is a candidate result of Q if it covers all the keywords in Q by $Q \subseteq \bigcup_{p \in A} p.\text{kw}$. Let S be the set including all candi- dates of Q. The top-1 result A^m of Q is obtained by

$$A^m = \arg\min_{A \in S} r(A)$$

:

Similarly, a top-k NKS query retrieves the top-k candidates with the least diameter. If two candidates have equal diame- ters, then they are further ranked by their cardinality.

Although existing techniques using tree-based indexes [2], [7], [8], [9] suggest possible solutions to NKS queries on multi-dimensional datasets, the performance of these algorithms deteriorates sharply with the increase of size or dimensional- ity in datasets. Our empirical results show that these algo- rithms may take hours to terminate for a multi-dimensional dataset of millions of points. Therefore, there is a need for an efficient algorithm that scales with dataset dimension, and yields practical query efficiency on large datasets.

In this paper, we propose ProMiSH (short for Projection and Multi-Scale Hashing) to enable fast processing for NKS queries. In particular, we develop an exact ProMiSH (referred to as ProMiSH-E) that always retrieves the opti- mal top-k results, and an approximate ProMiSH (referred to as ProMiSH-A) that is more efficient in terms of time and space, and is able to obtain near-optimal results in practice. ProMiSH-E uses a set of hashtables and inverted indexes to perform a localized search. The hashing tech- nique is inspired by Locality Sensitive Hashing (LSH) [10], which is a state-of-the-art method for nearest neighbor search in high-dimensional spaces. Unlike LSH-based methods that allow only approximate search with probabi- listic guarantees, the index structure in ProMiSH-E sup- ports accurate search. ProMiSH-E creates hashtables at multiple bin-widths, called index levels. A single round of search in a hashtable yields subsets of points that contain query results, and ProMiSH-E explores each subset using a fast pruning-based algorithm. ProMiSH-A is an approxi- mate variation of ProMiSH-E for better time and space effi- ciency. We evaluate the performance of ProMiSH on both real and synthetic datasets and employ state-of-the-art VbR^m-Tree [2] and CoSKQ [8] as baselines. The empirical results reveal that ProMiSH consistently outperforms the baseline algorithms with up to 60 times of speedup, and ProMiSH-A is up to 16 times faster than ProMiSH-E obtaining near-optimal results.

Our main contributions are summarized as follows. (1) We propose a novel multi-scale index for exact and approximate NKS query processing. (2) We develop effi- cient search algorithms that work with the multi-scale indexes for fast query processing. (3) We conduct exten- sive experimental studies to demonstrate the performance of the proposed techniques.

The paper is organized as follows. We start with the related work in Section 2. Next, we present the index struc- ture for exact search (ProMiSH-E) in Section 3, the exact search algorithm in Section 4, and its optimization techni- ques in Section 5. In addition, we introduce the approximate algorithm (ProMiSH-A) and provide an analysis for its approximation ratio in Section 6. The time and space com- plexity for ProMiSH are analyzed in Section 7. Experimental results are presented in Section 8. Finally, We conclude this paper with future work in Section 9. A glossary of the nota- tions is shown in Table 1.

2 Related Work

A variety of related queries have been studied in literature on text-rich spatial datasets.

Location-specific keyword queries on the web and in the GIS systems [11], [12], [13], [14] were earlier answered using a combination of R-Tree [15] and inverted index. Felipe et al. [16] developed IR^2-Tree to rank objects from spatial datasets based on a combination of their distances to the query locations and the relevance of their text descriptions to the query keywords. Cong et al. [17] integrated R-tree and inverted file to answer a query similar to Felipe et al. [16] using a different ranking function. Martins et al. [18] computed text relevancy and location proximity independently, and then combined the two ranking scores.

Cao et al. [7] and Long et al. [8] proposed algorithms to retrieve a group of spatial web objects such that the group's keywords cover the query's keywords and the objects in the group are nearest to the query location and have the lowest inter-object distances. Other related queries include aggre- gate nearest keyword search in spatial databases [19], top-k preferential query [20], top-k sites in a spatial data based on their influence on feature points [21], and optimal location queries [22], [23].

Our work is different from these techniques. First, existing works mainly focus on the type of queries

ISBN 978-93-88808-61-3

the coordinates of query points are known [7], [8]. Even though it is possible to make their cost functions same to the cost function in NKS queries, such tuning does not change their techniques. The proposed techniques use location information as an integral part to perform a best- first search on the IR-Tree, and query coordinates play a fundamental role in almost every step of the algorithms to prune the search space. Moreover, these techniques do not provide concrete guidelines on how to enable efficient processing for the type of queries where query coordi- nates are missing. Second, in multi- dimensional spaces, it is difficult for users to provide meaningful coordinates, and our work deals with another type of queries where users can only provide keywords as input. Without query coordinates, it is difficult to adapt existing techniques to our problem. Note that a simple reduction that treats the coordinates of each data point as possible query coordi- nates suffers poor scalability. Third, we develop a novel index structure based on random projection with hashing. Unlike tree-like indexes adopted in existing works, our index is less sensitive to the increase of dimensions and scales well with multi-dimensional data.

Another track of related works deal with *m*-closest keywords queries [9]. In [9], bR*-Tree is developed based on a R*-tree [24] that stores bitmaps and minimum bounding rectangles (MBRs) of keywords in every node along with points MBRs. The candidates are generated by the *apriori* algorithm [25]. Unwanted candidates are pruned based on the distances between MBRs of points or keywords and the best found diameter. However, the pruning techniques become ineffective with an increase in the dataset dimension as there is a large overlap between MBRs due to the curse of dimensionality. This leads to an exponential number of candidates and large query times. A poor estimation of starting diameter further worsens the performance of their algorithm. bR*-Tree also suffers from a high storage cost; therefore, Zhang et al. modified bR*-Tree to create Virtual bR*- Tree [2] in memory at run time. Virtual bR*-Tree is created from a pre-stored R*- Tree, which indexes all the points, and an inverted index which stores keyword information and path from the root node in R*-Tree for each point. Both bR*-Tree and Virtual bR*-Tree, are structurally similar, and use similar candi- date generation and pruning techniques. Therefore, Virtual bR*-Tree shares similar performance weaknesses as bR*-Tree.

Tree-based indexes, such as R-Tree [15] and M-Tree [26], have been extensively investigated for near- est neighbor search in high-dimensional spaces. These indexes fail to scale to dimensions greater than 10 because of the curse of dimensionality [27]. Random projection [28] with hashing [10], [29], [30], [31], [32] has come to be the state-of- the-art method for nearest neighbor search in high- dimensional datasets. Datar et al. [10] used random vectors constructed from *p*-stable distributions to project points, computed hash keys for the points by splitting the line of projected values into disjoint bins, and then concatenated hash keys obtained for a point from *m* ran- dom vectors to create a final hash key for the point. Our

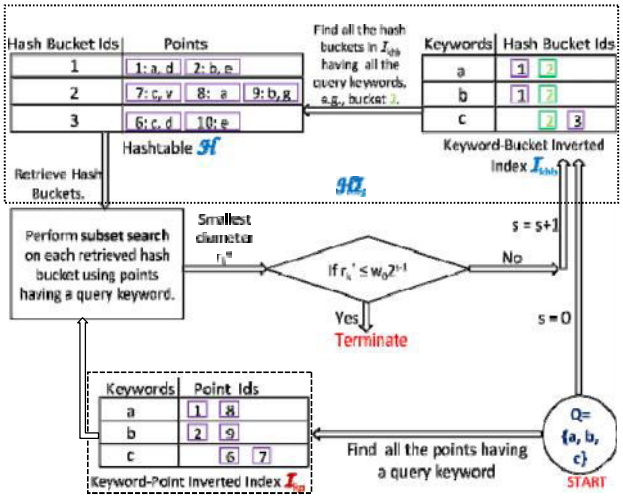


Fig. 2. Index structure and flow of execution of ProMiSH.

find the top-*k* tightest clusters that cover the input keyword set. Meanwhile, nearest neighbor queries usually require coordinate information for queries, which makes it difficult to develop an efficient method to solve NKS queries by existing techniques for nearest neighbor search. In addition, multi-way distance joins for a set of multi-dimensional datasets have been studied in [33], [34]. Tree-based index is adopted, but suffers poor scalability with respect to the dimension of the dataset. Furthermore, it is not straightforward to adapt these algorithms since every query requires a multi-way distance join only on a subset of the points of each dataset.

3 Index Structure for Exact ProMiSH

We start with the index for exact ProMiSH (ProMiSH-E). This index consists of two main components.

Inverted Index $\mathbf{I}_{kp}$. The first component is an inverted index referred to as $\mathbf{I}_{kp}$. In $\mathbf{I}_{kp}$, we treat keywords as keys, and each keyword points to a set of data points that are associated with the keyword. Let D be a set of data points and V be a dictionary that contains all the keywords appear- ing in D. We build $\mathbf{I}_{kp}$ for D as follows. (1) For each $v \in V$, we create a key entry in $\mathbf{I}_{kp}$, and this key entry points to a set of data points $D_v = \{o \in D \mid v \in o.\text{doc}\}$ (i.e., a set includes all data points in D that contain keyword *v*). (2) We repeat

(1) until all the keywords in V are processed. In Fig. 2, an example for $\mathbf{I}_{kp}$ is shown in the dashed rectangle at the bottom.

Hashtable-Inverted Index Pairs HI. The second component consists of multiple hashtables and inverted indexes referred to as HI. HI is controlled by three parameters: (1) (*Index level*) *L*, (2) (*Number of random unit vectors*) *m*, and (3) (*hashtable size*) *B*. All the three parameters are non-negative integers. Next, we describe how these three parameters con- trol the construction of HI.

In general, HI contains *L* hashtable-inverted index pairs, characterized by $\{(H^{0,s}, \mathbf{I}^{0,s}) \mid s \in \{0, 1, 2, \ldots, L-$

ISBN 978-93-88808-61-3

problem is different from nearest neighbor search. NKS queries provide no coordinate information, and aim to

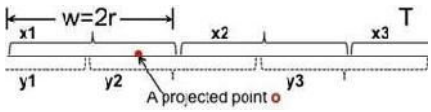


Fig. 3. The projection space (a segment) of a random unit vector is parti- tioned into overlapping bins of equal width: point *o* is included in bin *x*1 and *y*2.

First, given a set of *d*-dimensional data points D, we cre- ate hashtable $H^{\langle s \rangle}$ as follows.

1) We randomly sample *m* *d*-dimensional unit vectors $z_1; z_2; \ldots; z_m$ (i.e., $\|z_i\|_2 = 1$ for $i = 1; 2; \ldots; m$);
2) For each *o* ∈ D, we compute its projection on each of the unit vectors $o_{z_i} = o \cdot z_i$ where $i = 1; 2; \ldots; m$;
3) Let *pMax* be the maximum projected value for data points in D. For each z_i, we consider its projection space as a segment $[0; pMax]$, and partition the segment into $2^{L-s}+1$ overlapping bins, where each bin has width $w = w_0 2^s$ and is equally overlapped with two other bins as shown in Fig. 3. We conduct the projection space partition on all the *m* random unit vectors.
4) For each z_i and *o* ∈ D, since its projection space is partitioned into overlapping bins, o_{z_i} falls into two bins; therefore, we get two bin ids $\{b_1(o; z_i); b_2(o; z_i)\}$, and we can compute $b_1(o; z_i)$ and $b_2(o; z_i)$ as below,

$$b_1(o; z_i) = \left\lfloor \frac{o_{z_i}}{w} \right\rfloor \quad (1)$$

$$b_2(o; z_i) = \begin{cases} b_1(o; z_i) + 1 & \text{if } o_{z_i} - b_1(o; z_i) \cdot w \geq \frac{w}{2} \\ b_1(o; z_i) - 1 & \text{otherwise} \end{cases} \quad (2)$$

Fig. 3 shows an example where we partition the proj- ect space into overlapping bins $\{x1; x2; x3; y1; y2; y3\}$, and point *o* lies in bins *x*1 and *y*2.
5) For each *o* ∈ D, we generate its signatures based on the bins into which its projections on random unit vectors fall. With *m* random unit vectors, we obtain *m* pairs of bin ids for each data point *o*. Next, we take a cartesian product over these *m* pairs of bin ids and generate 2^m signatures for each point *o*, where each signature $\{b_{i_1}(o; z_1); \ldots; b_{i_m}(o; z_m)\}$ contains a bin id from each of the *m* pairs. For example, let z_1 and z_2 be two random unit vectors, and the bin ids for a point *o* be $\{x1; y1\}$ from z_1 and $\{x2; y2\}$ from z_2. We create four signatures as $\{x1; x2\}$, $\{x1; y2\}$, $\{y1; x2\}$, and $\{y1; y2\}$.
6) For each point *o* ∈ D, we hash it into 2^m buckets in $H^{\langle s \rangle}$ using its 2^m signatures. For each signature $\{b_{i_1}(o; z_1); \ldots; b_{i_m}(o; z_m)\}$, we convert it into a hashtable bucket id by a standard hash function, $\left(\sum_{j=1}^{m} b_{i_j} \cdot pr_j\right) \% B$, where *B* is hashtable size (i.e., the

to

where $H^{\langle s \rangle}$ and $I^{\langle s \rangle}$ are the *s*-th hashtable and inverted index, respectively.

each of which contains at least one data point *o* such that $v \in s(o)$. Fig. 2 demonstrates an example about HI with one pair of hashtable and inverted index shown in the dotted rectangle.

In next section, we show how to conduct exact search using ProMiSH-E.

4 Search Algorithm for ProMiSH-E

In this section, we present the search algorithms in ProM- iSH-E that finds top-*k* results for NKS queries. First, we introduce two lemmas that guarantee ProMiSH-E always retrieves the optimal top-*k* results. Then, we describe the details in ProMiSH-E. We start with some theoretic results for ProMiSH-E.

Lemma 1. *Let R^d be a d-dimensional euclidean space and z be a random unit vector uniformly sampled from R^d such that* $\|z\|_2 = 1$. *For any two points $o_1 \in R^d$ and $o_2 \in R^d$, we have* $\|o_1 - o_2\|_2 \leq \|o_{1z} - o_{2z}\|_2$, *where o_{1z} and o_{2z} are the projec- tion of o_1 and o_2 on z, respectively.*

Proof. Since Euclidean space with dot product is an inner product space, we have

$$\|o_{1z} - o_{2z}\|_2 = (o_1 - o_2) \cdot z \|_2 \leq \|z\|_2 \cdot \|o_1 - o_2\|_2 = \|o_1 - o_2\|_2; \text{ since } \|z\|_2 = 1:$$

The inequality follows the Cauchy-Schwarz inequality. □

Lemma 2. *Given $A = \{o_1; \ldots; o_n\} \subset R^d$ with diameter r is pro- jected onto a d-dimensional random unit vector z and the pro- jection space of z is partitioned into overlapping bins with equal width w, there exists at least one bin containing all the points in A if $w \leq 2r$.*

Proof. According to Lemma 1 and the definition of diameter, we have $\forall o_i, o_j \in A; |o_{iz} - o_{jz}| \leq \|o_i - o_j\|_2 \leq r$. Thus, we can further derive $\max(o_{1z}; \ldots; o_{nz}) - \min(o_{1z}; \ldots; o_{nz}) \leq r$. Since the projection space of *z* is partitioned into overlap- ping bins of width $w \leq 2r$, it follows from the construction that any line segment of width *r* is fully contained in one of the bins as shown in Fig. 3. Hence, all the points in *A* will fall into the same bin. □

We use an example to show how Lemma 2 guarantees the retrieval of the optimal top-1 results. Given a query *Q*, we assume the diameter of its top-1 result is *r*. We project all the data points in D on a unit random vector and parti- tion the projected values into overlapping bins of bin-width $w \leq 2r$. If we perform a search in each of the bins indepen- dently, then Lemma 2 guarantees that the top-1 result of query *Q* will be found in one of the bins. Based on Lemma 1 and Lemma 2, we propose ProMiSH-E as shown in Fig. 2. A search starts with the HI structure

ISBN 978-93-88808-61-3

number of buckets in $H^{\vartheta,P}$) and pr_j is a random prime number.

Second, given a dictionary V and hashtable $H^{\vartheta,P}$, we cre- ate the inverted index $\mathbf{I}^{\vartheta,\flat}$. In this inverted index, keys are still keywords. For each v 2 V_{khb} v points to a set of buckets,

at index level s ¼ 0. ProMiSH-E finds the buckets in hasht- able $H^{\vartheta,P}$, each of which contains all the query keywords by inverted index $\mathbf{I}^{\vartheta,\flat}$. Then, ProMiSH-E explores each selected bucket using an efficient pruning based technique to generate results. ProMiSH-E terminates after exploring

HI structure at the smallest index level s such that all the top-k results have been found.

the bucket from the hashtable H, and filters these points using bitset BS to get a subset of points F' in steps (17-22).
Subset F' contains only those points which are taggedwith

ISBN 978-93-88808-61-3

Algorithm 1. ProMiSH-E

In: Q: query keywords; k: number of top results
In: w_0: initial bin-width
1: PQ ½e½]; þ1Þ]: priority queue of top-k results 2: HC: hashtable to check duplicate candidates
3: BS : bitset to track points having a query keyword 4: for all o 2 [$_{8v_Q 2Q}$ I$_{kp}$½v_Q] do
5: BS½o] true /* Find points having query keywords*/ 6: end for
7: for all s 2 f0; … ;L — 1g do
8: Get HI at s
9: E½] 0 /* List of hash buckets */ 10: for all v_Q 2 Q do 11: for all bId 2 I$_{khb}$½v_Q] do
12: E½bId] E½bId] þ 1
13: end for
14: end for
15: for all i 2 ð0; … ; $SizeOf$ðEÞÞ do
16: if E½i] ¼ $SizeOf$ðQÞ then
17: F^0 ; /* Obtain a subset of points */ 18: for all o 2 H½i] do
19: if BS½o] ¼ true then
20: F^0 F^0 [o
21: end if
22: end for
23: if checkDuplicateCand(F^0; HC) = falsethen 24: searchInSubset(F^0, PQ)
25: end if
26: end if
27: end for
28: /* Check termination condition */ 29: if PQ½k]:$r \leq w$ 2^{s-1} then
30: Return PQ
31: end if
32: end for
33: /* Perform search on D if algorithm has not terminated */ 34: for all o 2 D do
35: if BS½o] ¼ true then
36: F^0 F^0 [o

at least one query keyword and is explored further.

Subset F^0 is checked whether it has been explored earlier or not using *checkDuplicateCand* (Algorithm 2) in step 23. Since each point is hashed using 2^m signatures, duplicate subsets may be generated. If F^0 has not been explored earlier, then ProMiSH-E performs a search on it using *searchInSubset* (Algorithm 3) at step 24 (We discuss in this algorithm in detail in Section 5). Results are inserted into a priority queue PQ of size k. Each entry of PQ is a tuple containing a set of points and their diame- ter. PQ is initialized with k entries, each of whose set is empty and the diameter is þ1. Entries of PQ are ordered by their diameters, and entries with equal diameters are further ordered by their set sizes. A new result is inserted into PQ only if its diameter is smaller than the k-th small- est diameter in PQ. If ProMiSH-E does not terminate after exploring the HI structure at index level s, then the search proceeds to HI at index level ðs þ 1Þ.

Algorithm 2. CheckDuplicateCand

In: F^0: a subset; HC: hashtable of subsets 1: F^0 $sort$ðF^{0}Þ
2: $pr1$: list of prime numbers; $pr2$: list of primenumbers; 3: for all o 2 F^0 do
4: pr_1 randomSelectð$pr1$Þ; pr_2 randomSelectð$pr2$Þ 5: h_1 h_1 þ ðo × pr_1Þ; h_2 h_2 þ ðo × pr_2Þ
6: end for
7: h $h_1 h_2$;
8: if isEmptyðHC½h]Þ=false then
9: if elementWiseMatch(F^0; HC½h]) = true then
10: Return true;
11: end if
12: end if
13: HC½h].addðF^0Þ;
14: Return false;

ProMiSH-E terminates when the k-th smallest diameter

Algorithm 1 details the steps in ProMiSH-E. It maintains a bitset BS. For each v_Q 2 Q, ProMiSH-E retrieves the list of points corresponding to v_Q from I$_{kp}$ in step 4. For each point o in the retrieved list, ProMiSH-E marks the bit correspond- ing to o's identifier in BS as true in step 5. Thus, ProMiSH-E finds all the points in D which are tagged with at least one query keyword. Next, the search continues in the HI struc- tures, beginning at s ¼ 0. For any given index level s, ProMiSH-E works with H$^{ðs\text{Þ}}$ and I$^{ðs\text{Þ}}_{khb}$ in HI at step 8. ProMiSH-E retrieves all the lists of hash bucket ids corresponding to keywords in Q from the inverted index I$^{ð0\text{Þ}}_{kp}$ at steps (10-11). An intersection of these lists yields a set of hash buckets each of which contains all the query keywords in steps (12-16) (e.g., In Fig. 2, this intersection yields the bucket id 2). For each selected hash bucket, ProMiSH-E retrieves all the points in

ISBN 978-93-88808-61-3

Lemma 2 guarantees that all the possible candidates are fully contained in one of the bins of the hashtable, and therefore, must have been explored. If ProMiSH-E fails to terminate after exploring HI at all the index levels s 2 f0; … L — 1g, then it performs a search in the complete dataset D during steps(34-39).

Algorithm *checkDuplicateCand* (Algorithm 2) uses a hashtable *HC* to check duplicates for a subset $F^{\emptyset}$. Points in F^{0} are sorted by their identifiers. Two separate standard hash functions are applied to the identifiers of the points in the sorted order to generate two hash values in steps (2-6). Both of the hash values are concatenated to get a hash key h for the subset F^{0} in step 7. The use of multiple hash func- tions helps to reduce hash collisions. If *HC* already has alist of subsets at h, then an element-wise match of F^{0} is per- formed with each subset in the list in steps (8-9). Otherwise, F^{0} is stored in *HC* using key h in step 13.

As shown in Algorithm 1, the efficiency of ProMiSH- E highly depends on an efficient search algorithm that finds top-k results from a subset of data points. In next sec- tion, we propose a search algorithm that provides such efficiency.

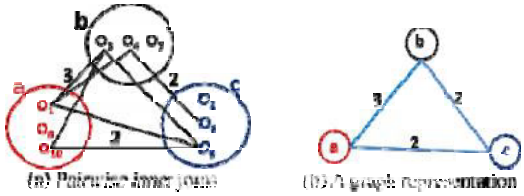


Fig. 4. (a) a, b, and c are groups of points of a subset F^{0} obtained for a query Q=fa; b; cg. A point o in a group g is joined to a point o' in another group g' if$\|o - o'\| \leq r_k$. The groups in the order {a, c, b} generates the least number of candidates by a multi-way join. (b) A graph of pairwise inner joins. Each group is a node in the graph. The weight of an edge is the num- ber of point pairs obtained by an inner join of the corresponding groups.

5 Search in a Subset of Data Points

We present an algorithm for finding top-k clusters tightest random. Let the edge ac, with weight 2, be selected in Fig. 4b. This forms the ordered set ða — cÞ. The next edge to be selected is the least weight edge such that at least one of its nodes is not included in the ordered set. Edge cb, with weight 2, is pickednextin Fig. 4b. Now the orderedsetis ða — c — bÞ. This process terminates when all the nodes are included in the set. ða — c — bÞ gives the orderingof the groups.

Algorithm 3 shows how the groups are ordered. The kth smallest diameter r_k is retrieved form the priority queue *PQ* in step 1. For a given subset F^{0} and a query Q, all the points are grouped using query keywords in steps (2-5). A pair- wise inner join of the groups is performed in steps (6-18). An adjacency list *AL* stores the distance between points which satisfy the distance threshold r_k. An adjacency list *M* stores the count of point pairs obtained for each pair of groups by the inner join. A greedy algorithm finds the order of the groups in steps (19-30). It repeatedly removes an edge with the smallest weight from *M* till all the groups are included in the order set *curOrder*. Finally, groups are sorted using *curOrder* in step 30. in a subset of points. A subset is obtained from a hasht-

We explain a multi-way distance join with an example. A multi-way distance join of q groups $\mathbf{f}g_1; \ldots; g_q\mathbf{g}$ finds all the tuples $\mathbf{f}o_{1;i}; \ldots; o_{;j}; o_{y;k}; \ldots; o_{q;l}\mathbf{g}$ such that 8x; y: $o_{x;j}$ 2 g_x , $o_{y;k}$ 2 g_y; and jj$o_{x;j}$ — $o_{y;k}$jj$_2 \leq r_k$. Fig. 4a shows groups $\mathbf{f}a, b, c\mathbf{g}$ of points obtained for a query Q ¼ $\mathbf{f}a, b, c\mathbf{g}$ from a subset $F^{\emptyset}$. We show an edge between a pair of points of two groups if the distance between the points is at most r_k, e.g., an edge between point o_1 in group a and point o_3 in group b. A multi- way distance join of these groups finds tuples $\mathbf{f}o_1, o_3, o_9\mathbf{g}$ and $\mathbf{f}o_{10}, o_3, o_9$ g. Each tuple obtained by a multi-way join is a promising candidate for a query.

Group Ordering

A suitable ordering ofthe groupsleads to anefficient candi- date exploration by a multi-way distance join. We first per- form a pairwise inner joins of the groups with distance threshold r_k. In inner join, a pair of points from two groups are joined only if the distance between them is at most r_k. Fig. 4a shows such a pairwise inner joins of the groups $\mathbf{f}a, b, c\mathbf{g}$. We see from Fig. 4a that a multi-way distance join in the order $\mathbf{f}a, b, c\mathbf{g}$ explores 2 true candidates $\mathbf{ff}o_1, o_3, o_9\mathbf{g}, \mathbf{f}o_{10}, o_3, o_9\mathbf{gg}$ and a false candidate $\mathbf{f}o_1, o_4, o_6\mathbf{g}$. A multi-way distance join in the order $\mathbf{f}a, c, b\mathbf{g}$ explores the least number of candidates 2. Therefore, a proper ordering of the groups leads to an effective pruning of false candidates. Optimal orderingofgroupsfortheleast numberofcandidates gener- ation is NP-hard [35].

We propose a greedy approach to find the ordering of groups. We explain the algorithm with a graph in Fig. 4b. Groups $\mathbf{f}a, b, c\mathbf{g}$ are nodes in the graph. The weight of an edge is the count of point pairs obtained by an inner join of the corresponding groups. The greedy method starts by selecting an edge having the least weight. If there are multiple edges with the same weight, then an edge is s lected at

Algorithm 3. SearchInSubset

In: $F^{\emptyset}$: subset of points; Q: query keywords; q: query size
In: *PQ*: priority queue of top-k results
1: r_k *PQ*.k]:r /* kth smallest diameter */
2: *SL* ½ðv;½]Þ]: list of lists to store groups per query keyword

ISBN 978-93-88808-61-3

```
for all v 2 Q do
    SL[v] ← {o ∈ F[i] : o is tagged with v} /* form groups */
end for
/* Pairwise inner joins of the groups*/
AL: adjacency list to store distances between points
M: adjacency list to store count of pairs between groups
for all (v_i, v_j) ∈ Q such that i ≤ q, j ≤ q, i < j do
    for all o ∈ SL[v_i] do
        for all o' ∈ SL[v_j] do
            if ||o − o'||_2 ≤ r_k then
                AL[o, o'] ← ||o − o'||_2
                M[v_i, v_j] ← M[v_i, v_j] + 1
            end if
        end for
    end for
end for
/* Order groups by a greedy approach */ curOrder = []
while Q ≠ ∅ do
    (v_i, v_j) ← removeSmallestEdge(M)
    if v_i ∉ curOrder then
        curOrder.append(v_i); Q ← Q − v_i
    end if
    if v_j ∉ curOrder then
        curOrder.append(v_j); Q ← Q − v_j
    end if
end while
sort(SL, curOrder) /* order groups */
findCandidates(q, AL, PQ, Idx, SL, curSet, curSetr, r_k)
```

Nested Loops with Pruning

We perform a multi-way distance join of the groups by nested loops. For example, consider the set of points in

Fig. 4. Each point $o_{a;i}$ of group a is checked against each point $o_{b;j}$ of group b for the distance predicate, i.e., $\|o_{a;i} - o_{b;j}\|_2 \leq r_k$. If a pair $(o_{a;i}, o_{b;j})$ satisfies the distance predicate, then it forms a tuple of size 2. Next, this tuple is checked against each point of group c. If a point $o_{c;k}$ satisfies the distance predicate with both the points $o_{a;i}$ and $o_{b;j}$, then a tuple $(o_{a;i}, o_{b;j}, o_{c;k})$ of size 3 is generated. Each intermedi- ate tuple generated by nested loops satisfies the property that the distance between every pair of its points is at most r_k. This property effectively prunes false tuples very early in the join process and helps to gain high efficiency. A can- didate is found when a tuple of size q is generated. If a can- didate having a diameter smaller than the current value of r_k is found, then the priority queue PQ and the value of r_k are updated. The new value of r_k is used as distance thresh- old for future iterations of nestedloops.

Algorithm 4. findCandidates

```
In: q: query size; SL: list of groups
In: AL: adjacency list of distances between points
In: PQ: priority queue of top-k results
In: Idx: group index in SL
In: curSet: an intermediate tuple
In: curSetr: an intermediate tuple's diameter 1:          if Idx ≤ q then
 2:     for all o 2 SL[Idx] do
 3:       if AL[curSet[Idx-1], o] ≤ r_k then
 4:         newCurSetr curSetr
 5:         for all o' 2 curSetdo
 6:           dist    AL[o, o']
 7:           if dist ≤ r_kthen
 8:             flag  true
 9:             if newCurSetr < dist  then
10:               newCurSetr dist
11:             end if
12:           else
13:             flag false; break;
14:           end if
15:         end for
16:         if flag = true then
17:           newCurSet  curSet.append(o)
18:           r_k findCandidates(q, AL, PQ, Idx þ 1; SL, newCurSet, newCurSetr, r_k)
19:         else
20:           Continue;
21:         end if
22:       end if
23:     end for
24:     return r_k
25:   else
26:     if checkDuplicateAnswers(curSet, PQ) = true then
27:       return r_k
28:     else
```

An intermediate tuple *curSet* is checked against each point of group $SL[Idx]$ in steps (2-23). First, it is determined using AL whether the distance between the last point in *curSet* and a point o in $SL[Idx]$ is at most r_k in step 3. Then, the point o is checked against each point in *curSet* for the dis- tance predicate in steps (5-15). The diameter of *curSet* is updated in steps (9-11). If a point o satisfies the distance predicate with each point of *curSet*, then a new tuple *newCurSet* is formed in step 17 by appending o to *curSet*. Next, a recursive call is made to *findCandidates* on the next group $SL[Idx + 1]$ with *newCurSet* and *newCurSetr*. A can- didate is found if *curSet* has a point from every group. A result is inserted into PQ after checking for duplicates in steps (26-33). A duplicate check is done by a sequential match with the results in PQ. For a large value of k, a method similar to Algorithm 2 can be used for a duplicate check. If a new result gets inserted into PQ, then the value of r_k is updated in step 18.

6 Approximate Search: ProMiSH-A

In this section, we discuss the approximate version of ProM- iSH referred to as ProMiSH-A. We start with the algorithm description of ProMiSH-A, and then analyze its approxima- tion quality.

Algorithm overview. In general, ProMiSH-A is more time and space efficient than ProMiSH-E, and is able to obtain near-optimal results in practice. The index structure and the search method of ProMiSH-A are similar to ProMiSH-E; therefore, we only describe the differences between them.

The index structure of ProMiSH-A differs from ProM- iSH-E in the way of partitioning projection space of random unit

vectors. ProMiSH-A partitions projection space into non-overlapping bins of equal width, unlike ProMiSH-E which partitions projection space into overlapping bins. Therefore, each data point *o* gets one bin id from a random unit vector *z* in ProMiSH-A. Only one signature is generated for each point *o* by the concatenation of its bin ids obtained from each of the *m* random unit vectors. Each point is hashed into a hashtable using its signature.

The search algorithm in ProMiSH-A differs from ProM- iSH-E in the termination condition. ProMiSH-A checks for a termination condition after fully exploring a hashtable at a given index level: It terminates if it has *k* entries with non- empty data point sets in its priority queue *PQ*.

Approximation quality analysis. In the following, we ana- lyze the approximation quality for the top-1 result returned by ProMiSH-A. In particular, we use approximation ratio r ≤ 1 as the metric to evaluate approximation quality. This ratio is defined as the ratio of the diameter of the result reported by ProMiSH-A *r* to the diameter of the optimal result r^m: r ¼ $r=r^m$. Let D be a *d*-dimensional dataset and

29: if *curSetr* ≺ *PQ*[*k*]:*r* then 30: *PQ.Insert*([*curSet*, *curSetr*]) 31:
32: end if
33: end if
34: end if
return *PQ*[*k*]:*r*

We find results by nested loops as shown in Algorithm 4 that each data point in D has *t* keywords, and each keyword is independently sampled by a uniform distribution over a dictionary V with *U* unique keywords. We define $f(v) = 1 - (1 - 1/U)^t$ as the probability that a data point has keyword *v* 2 V. Thus, we can estimate the expected number of points that have keyword *v* as $E[N_v] = Nf(v)$. To this end, the expected number of candidates for query *Q* in D is estimated by N_q

Let $g(r)$ be the probability that a candidate has a diameter no more than *r*. Then, the expected number of candidates for query *Q* with diameter no more than *r* is estimated by

$$N_r = g(r) N^q \prod_{i=1}^{q} f(v_i).$$

We index data points in D by ProMiSH-A, where each data point is projected onto *m* random unit vectors. The pro- jection space of each random unit vector is partitioned into non-overlapping bins of equal width *w*. Let $Pr(A; r \mid w)$ be the conditional probability for random unit vectors that a candidate *A* of query *Q* having diameter *r* is fully contained within a bin with width *w*. For *m* independent unitrandom vectors, the joint probability that a candidate *A* is contained in a bin in each of the *m* vectors is $Pr(A; r \mid m)^m$, and the probability that no candidate of diameter *r* is retrieved by ProMiSH-A from the hashtable, created using *m* unit ran- dom vectors, is $(1 - Pr(A; r \mid w)^m)^{N_r}$. Let the diameter of the top-1 result of query *Q* be r^*. Then, the probability $P(r^*)$ of at least one candidate of any diameter *r*, where $r^* \le r \le r^l$, being retrieved by ProMiSH-A is given by

$$P(r^*) = 1 - \prod_{r=r^*}^{r^l} (1 - Pr(A; r \mid w)^m)^{N_r} : \quad (3)$$

For a given constant *Z*, $0 \le Z \le 1$, we can compute the smallest value of r^l using Equation (3) such that $Z \le P(r^*)$. The value $\bar{r} = r^l / r^*$ gives the approximation ratio of the results returned by ProMiSH-A with the probability *Z*.

We empirically computed r for queries of three key- words for different values of *Z* using this model. We used a 32-dimensional real dataset having one million points described in Section 8 for our investigation, and computed the values of N_r and $Pr(A; r \mid w)^2$, where we use two random unit vectors with bin-width of *w* ¼ 100. In this way, we obtained the approximation ratio bound of r ¼ 1:4 and r ¼ 1:5 for *Z* ¼ 0:8 and *Z* ¼ 0:95, respectively.

in a bucket *b* among *B* buckets. Suppose ProMiSH-E applies *m* random unit vectors. Since ProMiSH-E generates 2^m sig- natures for each data point, the expectation of $N_{b;v}$ under uniformity assumptions is estimated as below,

$$E[N_{b;v}] = \frac{2^m}{B} E[N_v]:$$

Searching a bucket *b* in $H^{(s)}$ includes inner group joins and nested loops. Let *q* be the number of keywords in a query. First, inner group joins for *d*-dimensional data points are computed in $O(d(E[N_{b;v}])^2 + q E[N_{b;v}])^2)$. Second, nested loops are computed in $O((E[N_{b;v}])^q)$. Thus, the total complexity of searching a bucket *b* is $O(d(E[N_{b;v}])^2 + (qE[N_{b;v}])^2 + (E[N_{b;v}])^q)$. In the worst case, we may need to check all the buckets at all scales; therefore, the overall complexity is $O(dLB(E[N_{b;v}])^2 + LB(qE[N_{b;v}])^2 + LB(E[N_{b;v}])^q)$.

Time complexity of ProMiSH-A. Let *L* be the index level applied in the index structure of ProMiSH-A, $H^{(s)}$ be the hashtable at scale *s* 2 f0; 1; ... ;*L* — 1g, *B* be hashtable size, and $N^0_{b;v}$ be the number of data points with keyword *v* lying in a bucket *b*. Since ProMiSH-A only generates one signature per data point, the expectation of $N^0_{b;v}$ under uniformity assumptions is estimated as $\frac{E[N_v]}{B}$

7 Complexity Analysis of ProMiSH

In this section, we first analyze the query time complexity and index space complexity in ProMiSH. Then we discuss how ProMiSH prunes the search space.

ISBN 978-93-88808-61-3

Query Time Complexity

Given a set of *d*-dimensional data points D, we assume data points are uniformly distributed in the buckets of a hasht- able, and keywords of each data point are uniformly sam- pled from the dictionary.

Suppose D has *N* data points, each data point has *t* key- words, and the keywords are sampled from a dictionary of *U* unique keywords. Let N_v be the number of data points with keyword *v*. The expectation of N_v is computed as follows,

$$E[N_v] = \frac{N}{B}.$$

Similarly, we can derive the overall complexity of ProMiSH- A is $O(dLB(E[N_{b,v}])^2 + LB(qE[N_{b,v}])^2 + LB(E[N_b])^2)$.

Index Space Complexity

Let *N* be the number of data points to index, *d* be the dimen- sion of data points, *t* be the number of keywords per data point, *U* be dictionary size, *m* be the number of random unit vectors for point projection, *L* be index level, and *B* be hashtable size.

Space complexity of ProMiSH-E. The Indexes of ProMiSH- E includes keyword-point inverted index $\mathbf{I}_{kp}$, hashtable H, and keyword-bucket inverted index $\mathbf{I}_{khb}$. First, we need $O(tN + U)$ space for $\mathbf{I}_{kp}$. Second, in the worst case, we need to include $O(2^m NL + BL)$ points in H. Finally, $\mathbf{I}_{khb}$ takes $O(UBL)$ space. Thus, at index level *L*, the overall space com- plexity is $O(tN + 2^m NL + UBL)$.

Space complexity of ProMiSH-A. The Indexes of ProMiSH-A also includes keyword-point inverted index $\mathbf{I}_{kp}$, hasht- able H, and keyword-bucket inverted index $\mathbf{I}_{khb}$. Unlike ProMiSH-E, H in ProMiSH-A takes at most $O(NL + BL)$. Thus, at index level *L*, the overall space complexity is $O(tN + NL + UBL)$.

Pruning Intuition

Let D be a *d*-dimensional dataset of *N* data points, *U* be dic-

ISBN 978-93-88808-61-3

Time complexity of ProMiSH-E. Let L be the index level applied in the index structure of ProMiSH-E, $H^{\vartheta s Þ}$ be the hashtable at scale s 2 f0; 1; ... ;L — 1g, B be hashtable size, and $N_{b;v}$ be the number of data points with keyword $v$lying assume each data point is associated with only one keyword.

Suppose node set A^{m} c D with diameter r^{m} is the top-1 result for query Q. Let $f\vartheta vÞ$ denote the probability that a

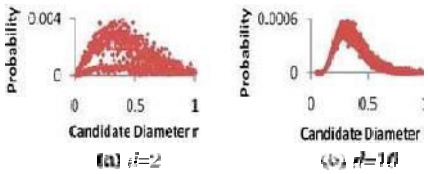


Fig. 5. Candidate diameter distributions for queries with three keywords over a two-dimensional and a 16-dimensional real datasets.

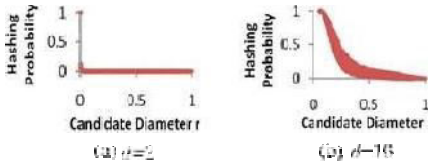


Fig. 6. Distributions of $Pr\vartheta A; rjwÞ^{2}$ for candidates of queries with three keywords over a two-dimensional and a 16-dimensional real datasets.

data point has keyword v and $g\vartheta rÞ$ denote the probability that a candidate of Q has diameter no more than r. Given query Q, the expected number of candidates N_Q and the expected number of candidates $N_{Q;r}$ with diameter no more than r are calculated as follows,

$$N_Q ¼ \prod_{i¼1} \vartheta f\vartheta v_i ÞNÞ; N_{Q;r} ¼ g\vartheta rÞN_Q:$$

We select all the points in D which contain at least one query keyword v_i, project these points on a random unit vector, and split the line of projected values into overlap- ping bins of equal width $w ¼ 2r^{m}$. Let $Pr\vartheta A; r$ j $wÞ$ be the con- ditional probability for random unit vectors that a candidate A with diameter r is fully contained within a bin of width w. For m independent random unit vectors, the probability that a candidate A is contained in a bin in each of the m vectors is $Pr\vartheta A; r$ j $wÞ^{m}$. Ideally, the expected num- ber of candidates explored by ProMiSH in a hashtable is

$$N_p ¼ \sum_{i¼0}^{m} Pr\vartheta A; r \text{ j } w_i Þ \quad N_Q:$$

We empirically measured keyword distribution $f\vartheta Þ$, $Pr\vartheta A; r$ j $wÞ^{m}$, and the ratio of N_p to N_Q by real datasets of one million data points with varied dimensions (more details about the dataset are described in Section 8).

Candidate diameter distributions and the distributions of $Pr\vartheta A; r$ j $wÞ^{2}$ are demonstrated in Figs. 5 and 6, where candi- date diameters are scaled to between 0 and 1. We make following observations. (1) Candidate diameters follow a heavy-tailed distribution, which suggests a large number of candidates have diameters much larger than r^{m}. (2) The distributions of $Pr\vartheta A; r$ j $wÞ^{2}$ decreases exponentially with can- didate diameter, which implies that the candidates with diameter larger than r^{m} have much smaller chance of falling in a bin and being probed by ProMiSH, compared with A^{m}. Therefore, most of candidates with diameters larger than r^{m} are effectively pruned out by ProMiSH using its index.

Table 2 presents the ratios of N_p to N_Q. Each ratio is com- puted as an average of 50 random queries. We observe that

TABLE 2
Ratios of the Expected Number of Candidates N_p to the Expected Number of Candidates N_Q

Dataset Dimension d	2	4	8	16	32
Percentage ratio (N^{p}) N_Q	0.007	0.3	5.8	22	47

TABLE 3
Statistics of Datasets Used in Experiments

ISBN 978-93-88808-61-3

Dataset	Dataset size N	Dictionary size U	Keywords per point t
Real-1	10;000	5;661	12
Real-2	30;000	6;753	13
Real-3	50;000	7;101	13
Real-4	70;000	7;902	14
Real-5	1;000;000	24;874	11

ProMiSH prunes more than 99 and 50 percent of false candi- dates for d ¼ 2 and d ¼ 32, respectively.

8 Experimental Results

In this section, we evaluate the effectiveness and efficiency of ProMiSH by both real and synthetic data.

Setup

Dataset. Our evaluation employs real and synthetic datasets. The real datasets are collected from photo-sharing web- sites. As discussed in Section 1, one of the applications in NKS queries is to find tight clusters of photos which contain all the keywords provided by a user in a photo-sharing social network. We crawl images with descriptive tags from Flickr,[2] and then these images are transformed into gray- scale. Let d be the desired dimensionality. We convert each image into a d- dimensional point by extracting its color his- togram, and associate each data point with a set of key- words that are derived from its tags. In total, we collect five datasets (referred to as Real-1, Real-2, Real-3, Real-4, and Real-5) with up to one million data points. Their statistics are shown in Table 3.

We also generate synthetic datasets to evaluate the scal- ability of ProMiSH. In particular, the data generation pro- cess is governed by the following parameters: (1) Dimension d specifies the dimensionality of each data point; (2) Dataset size N indicates the total number of multi- dimensional points in a synthetic dataset; (3) Keywords per point t suggests the number of keywords for each data point; and (4) Dictionary size U denotes the total number of keywords in a dataset. For each data point, its coordinate in each dimension is randomly sampled between 0 and 10;000, and its keyword is randomly selected following a uniform distribution. We create multiple synthetic datasets to inves- tigate how these parameters affect the performance of ProMiSH.

Query. We generate NKS queries for real and synthetic datasets. In general, the query generation process is con- trolled by two parameters: (1) Keywords per query q decides the number of keywords in each query; and (2)

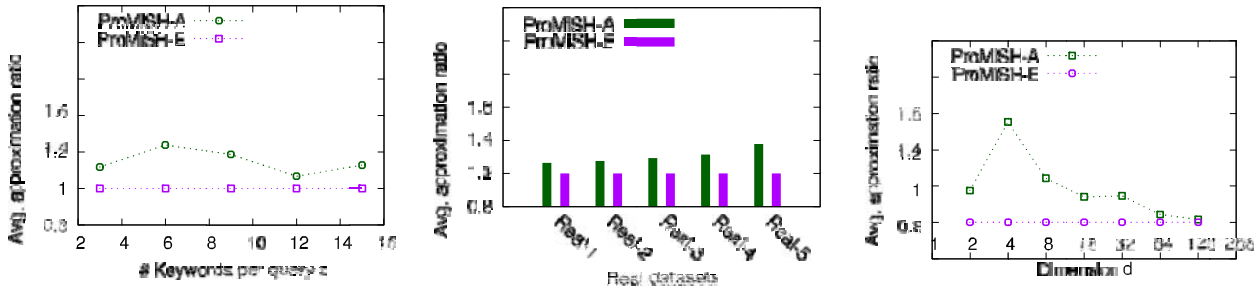


Fig. 7. Average approximation ratio of ProMiSH-A under different input real data: (left) varying the number of keywords per query q; (middle) varying real datasets; and (right) varying the number of dimensions in data points d.

Dictionary size U indicates the total number of keywords in a target dataset. For a real dataset, the probability that a key- word will be sampled in a query is proportional to the key- word's frequency in the dataset. For a synthetic dataset, a keyword of a query is randomly sampled following a uni- form distribution.

Implementation. In addition to the exact ProMiSH (ProM- iSH-E) and the approximate ProMiSH (ProMiSH-A), we also implement Virtual bR*-Tree (VbR^m-Tree) [2] and CoSKQ [7], [8] as baselines.

For VbR^m-Tree, we fix the leaf node size to be 1;000 entries and other node size to be 100 entries, as it demon- strate the best performance under this parameter setting.

CoSKQ is designed to handle the type of queries with query coordinates. To adapt CoSKQ to our problem, we transform an NKS query into a set of CoSKQ queries. Given a data point from a target dataset and an NKS query, we build a CoSKQ query by using the coordinates of the data point and the keywords in the NKS query. To ensure the correctness, if a dataset has N data points, we have to build N CoSKQ queries that enumerate all possible query coordinates.

All the algorithms are implemented in C++ with GCC 4.8.2, and all the experiments are conducted on a server with Ubuntu 14.04, powered by an Intel Core i7-2620M 2.7GHz CPU and 64 GB of RAM. Each experiment is

ISBN 978-93-88808-61-3

repeated 10 times, and their average results from 100 queries are presented.

Effectiveness

We apply real datasets to demonstrate the effectiveness of ProMiSH-A. In particular, we use the metric *approximation ratio* [30], [32] for evaluation. Let Q be an NKS query, r_i bethe ith smallest diameter from the top-k results returned by ProMiSH-A, and r^*_i be the ith smallest diameter returned by ProMiSH-E. The approximation ratio of ProMiSH-A with respect to Q is defined by $r(Q) = \frac{1}{k}\sum_{i=1}^{k} \frac{r_i}{r^*_i}$. It is easy to see $r(Q) \geq 1$; moreover, the smaller $r(Q)$ is, the better the algo- rithm will be with respect to Q. In the following, we report the average approximation ratio (AAR), which is the mean of the approximation ratios over all evaluated queries.

Fig. 7 shows the effectiveness of ProMiSH-A under dif- ferent input real data. In this set of experiments, the index parameters are fixed as $m = 4$, $L = 5$, and $B = 10{,}000$. We range the dataset among Real-1, Real-2, Real-3, Real-4, and Real-5 with Real-3 as the default dataset, the number of keywords per query q from 3 to 15 with 9 as the default q, and the number of dimensions for data points d from 2 to 128 with 16 as the default d. All the algorithms focus on top- 1 result. The results demonstrate that the AAR of ProMiSH- A is no more than 1.6 in all circumstances, and is no more than 1.2 in most cases.

Fig. 8 reports the effectiveness of ProMiSH-A under different index parameters over the real dataset Real-3. In this set of experiments, we fix the dimensions of the data points in Real-3 to be 16, and the number of keywords in queries to be 9. For index parameters, we vary the num- ber of random unit vectors m from 2 to 6 with 4 as the default m, index level L from 5 to 13 with 5 as the default m, and hashtable size B from 1,000 to 100,000 with 10,000 as the default B. All the algorithms focus on top-1 result. In general, the AAR of ProMiSH-A is no more than 1.3 in all circumstances. Moreover, we observe the following trends: (1) when m increases, the AAR increases; (2) when L increases, the AAR decreases; and (3) when B increases, the AAR increases.

Fig. 8. Average approximation ratio of ProMiSH-A under different index parameters over Real-3: (left) varying the number of random

ISBN 978-93-88808-61-3

Fig. 9 reports the effectiveness of ProMiSH-A in different top-k search over the real dataset Real-3. In this experiment, the input parameters are fixed as d ¼ 16 and q ¼ 9; and the index parameters are fixed as m ¼ 4, L ¼ 5, and B ¼ 10;000. As k is varied from 1 to 9, the AAR of ProMiSH-A is no more than 1:2.

In summary, Figs. 7, 8, and 9 consistently suggest the high effectiveness of ProMiSH-A.

Efficiency

We employ response time as the metric to evaluate the effi- ciency of different algorithms. Given a set of queries, the response time of an algorithm is defined as the average amount of time that the algorithm spends in processing one query.

Fig. 10 presents the response time of ProMiSH-E, ProM- iSH-A, and VbR^m-Tree under different input real data. In this set of experiments, the index parameters are fixed as m ¼ 4, L ¼ 5, and B ¼ 10;000. We range the dataset among Real- 1, Real-2, Real-3, Real-4, and Real-5 with Real-3 as the default dataset, the number of keywords per query q from 3 to 15 with 9 as the default q, and the number of dimensions for data points d from 2 to 128 with 16 as the default d. All the algorithms focus on top-1 result. Note that the result of CoSKQ based method is not shown in Fig. 10, as it cannot finish this experiment within one day. We make the follow- ing observations based the results. (1) As the number of key- words per query q increases, the response time of all algorithms increases. Compared with VbR^m-Tree, ProMiSH- E and ProMiSH-A are up to 30 and 60 times faster, respec- tively. (2) In all real datasets, ProMiSH-E and ProMiSH-A consistently outperform VbR^m-Tree with up to 18 and 25 times of speedup, respectively. Moreover, VbR^m-Tree can- not process the workload for Real-5 within one day. (3) When d is ranged from 2 to 128, ProMiSH-E and ProMiSH- A can finish the computation within one second. As one has to transform an NKS query into thousands of CoSKQ queries for the correctness and evaluate them all, CoSKQ based method processes a query in 2 to 10 seconds (not shown) even when d is 2 or 4, which is up to 100 times slower than our methods. In terms of VbR^m-Tree, it finishes the computation in more than one minute for d ¼ 32 (not shown), but cannot finish this experiment within one day. (4) ProMiSH-A outperforms ProMiSH-E with up to 16 times of speedup.

Fig. 11 shows the efficiency of ProMiSH-E and ProM- iSH-A under different index parameters over the real dataset Real-3. In this set of experiments, we fix the dimensions of the data points in Real-3 to be 16, and the number of keywords in queries to be 9. For index param- eters, we vary the number of random unit vectors m from 2 to 6 with 4 as the default m, index level L from 5 to 13 with 5 as the default m, and hashtable size B from 1;000 to 100;000 with 10;000 as the default B. All the algorithms focus on top-1 result. From the results, we observe that (1) m ¼ 4 empirically renders the best response time for both ProMiSH-E and ProMiSH-A; (2) as L increases, the response time of both algorithms decreases; and (3) hasht- able size has minor influence on the response time of the two algorithms.

Fig. 12 reports the response time of the algorithms on searching top-k results over Real-3. In this experiment, the input parameters are fixed as d ¼ 16 and q ¼ 9; and the index parameters are fixed as m ¼ 4, L ¼ 5, and B ¼ 10;000. Note that VbR^m-Tree and the CoSKQ based method are excluded from this experiment since they mainly support top-1 search. The results indicate that (1) as k increases, the response time of both algorithms increases; and (2) ProMiSH-A is consistently faster than ProMiSH-E.

Fig. 13 presents the response time of the algorithms under different synthetic data. In this set of experiments, the index parameters are fixed as m ¼ 4, L ¼ 5, and B ¼ 10 000, and we apply 6 parameters to control syn- thetic data generation: (1) the number of keywords per query q, ranging from 3 to 15 with 9 as the default q; (2) dataset size N, ranging from 10;000 to 10;000;000 with 1;000;000 as the default N; (3) data point dimension d, ranging from 2 to 128 with 16 as the default d; (4) the number of keywords per data point t, ranging from 1 to 16 with 4 as the default t; (5) dictionary size U, ranging from 100 to 10;000 with 1;000 as the default U; and (6) the k in top-k search, ranging from 1 to 9, with 1 as the default k. Note that the results of VbR^m-Tree and the CoSKQ based method are not shown here since they can- not finish this experiment within one day. We draw the following observations based on the results. (1) As q, N, d, t, or k increases, the response time of ProMiSH-E and ProMiSH-A increases. (2) As U increases, the response time of both algorithms decreases. (3) ProMiSH-A can process one query in 2 minutes in all cases, while ProM- iSH-E processes one query in 20 minutes in average. (4) ProMiSH-A outperforms ProMiSH-E in terms of response time with up to 100 times of speedup.

Fig. 14 reports the response time of ProMiSH-E and ProMiSH-A under different index parameters over syn- thetic data. In this set of experiments, the synthetic data are generated with a parameter setting q ¼ 9, N ¼ 1;000;000, d ¼ 16, t ¼ 4, U ¼ 1;000, and k ¼ 1. For index parameters, we range the number of random unit vectors m from 2 to 6 with 4 as the default m, index level L from 5 to 13 with 5 as the default L, and hashtable size B from 1;000 to 100;000 with 10;000 as the default B. We observe that (1) m ¼ 4 ren- ders the best response time for ProMiSH-E; (2) as L increases, ProMiSH-A obtains significant improvement in terms of efficiency; (3) hashtable size B has minor influence on both algorithms'

response time; and (4) ProMiSH-A is up to 200 times faster than ProMiSH-E.

Index Efficiency

We use memory usage and indexing time as the metrics to evaluate the index size for ProMiSH-E and ProMiSH-A. In particular, Indexing time indicates the amount of time used to build ProMiSH variants.

Fig. 15 presents the memory usage and indexing time of ProMiSH-E and ProMiSH-A under different input real

data. In this set of experiments, the index parameters are fixed as m ¼ 4, L ¼ 5, and B ¼ 10;000. We vary the num- ber of dimensions in data points m from 2 to 128 with 16 as the default m, and the datasets among Real-1, Real-2, Real-3, Real-4, and Real-5 with Real-3 as the default data- set. From the results, we make the following observations.

(1) Memory usage grows slowly in both ProMiSH-E and ProMiSH-A when the number of dimensions in data points increases. (2) ProMiSH-A is more efficient than ProMiSH-E in terms of memory usage and indexing time: it takes 80 percent less memory and 90 percent less time, and is able to obtain near-optimal results as shown in Fig. 7. (3) Over all cases, the memory usage ratio of ProM- iSH to raw data is no more than 13:4 for ProMiSH-E and no more than 2:4 for ProMiSH-A.

Summary

We summarize the experimental results as follows. First, ProMiSH-E and ProMiSH-A consistently outperform the baseline methods in terms of efficiency with up to 60 times of speedup. Second, ProMiSH-A is up to 16 times faster than ProMiSH-E, and can obtain near-optimal results. Third, ProMiSH-A is more space-efficient: compared with ProM- iSH-E, it takes 80 percent less memory and 90 percent less indexing time.

9 Conclusions and Future Work

In this paper, we proposed solutions to the problem of top-k nearest keyword set search n multi-dimensional datasets. We proposed a novel index called ProMiSH based on ran- dom projections and hashing. Based on this index, we developed ProMiSH-E that finds an optimal subset of points and ProMiSH-A that searches near-optimal results with bet-ter efficiency. Our empirical results show that ProMiSH is faster than state-of-the-art tree-based techniques, with mul- tiple orders of magnitude performance improvement. More- over, our techniques scale well with both real and synthetic datasets.

Ranking functions. In the future, we plan to explore other scoring schemes for ranking the result sets. In one scheme, we may assign weights to the keywords of a point by using techniques like tf-idf. Then, each group of points can be scored based on distance between points and weights of keywords. Furthermore, the criteria of a result containing all the keywords can be relaxed to generate results having only a subset of the query keywords.

Disk extension. We plan to explore the extension of ProM- iSH to disk. ProMiSH-E sequentially reads only required buckets from $\mathbf{I}_{kp}$ to find points containing at least one query keyword. Therefore, $\mathbf{I}_{kp}$ can be stored on disk using a direc- tory-file structure. We can create a directory for $\mathbf{I}_{kp}$. Each bucket of $\mathbf{I}_{kp}$ will be stored in a separate file named after its key in the directory. Moreover, ProMiSH-E sequentially probes HI data structures starting at the smallest scale to generate the candidate point ids for the subset search, andit reads only required buckets from the hashtable and the inverted index of a HI structure. Therefore, all the hasht- ables and the inverted indexes of HI can again be stored using a similar directory-file structure as $\mathbf{I}_{kp}$, and all the points in the dataset can be indexed into a B+-Tree [36] using their ids and stored on the disk. In this way, subset search can retrieve the points from the disk using B+-Tree for exploring the final set of results.

Acknowledgments

This research work is supported in part by the US National Science Foundation (NSF) under grant IIS- 1219254.

References

[1] W. Li and C. X. Chen, "Efficient data modeling and querying system for multi-dimensional spatial data," in *Proc. 16th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2008, pp. 58:1–58:4.

[2] D. Zhang, B. C. Ooi, and A. K. H. Tung, "Locating mapped resources in web 2.0," in *Proc. IEEE 26th Int. Conf. Data Eng.*, 2010, pp. 521–532.

[3] V. Singh, S. Venkatesha, and A. K. Singh, "Geo-clustering of images with missing geotags," in *Proc. IEEE Int. Conf. Granular Comput.*, 2010, pp. 420–425.

[4] V. Singh, A. Bhattacharya, and A. K. Singh, "Querying spatial patterns," in *Proc. 13th Int. Conf. Extending Database Technol.: Adv. Database Technol.*, 2010, pp. 418–429.

[5] J. Bourgain, "On lipschitz embedding of finite metric spaces in hilbert space," *Israel J. Math.*, vol. 52, pp. 46–52, 1985.

[6] H. He and A. K. Singh, "GraphRank: Statistical modeling and mining of significant subgraphs in the feature space," in *Proc. 6th Int. Conf. Data Mining*, 2006, pp. 885–890.

[7] X. Cao, G. Cong, C. S. Jensen, and B. C. Ooi, "Collective spatial keyword querying," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2011, pp. 373–384.

[8] C. Long, R. C.-W. Wong, K. Wang, and A. W.-C. Fu, "Collective spatial keyword queries: A distance owner-driven approach," *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2013, pp. 689–700.

[9] D. Zhang, Y. M. Chee, A. Mondal, A. K. H. Tung, and M. Kitsure- gawa, "Keyword search in spatial databases: Towards searching by document," in *Proc. IEEE 25th Int. Conf. Data Eng.*, 2009, pp. 688–699.

[10] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, "Locality- sensitive hashing scheme based on p-stable distributions," in *Proc. 20th Annu. Symp. Comput. Geometry*, 2004, pp. 253–262.

[11] Y. Zhou, X. Xie, C. Wang, Y. Gong, and W.-Y. Ma, "Hybrid index structures for location-based web search," in *Proc. 14th ACM Int. Conf. Inf. Knowl. Manage.*, 2005, pp.155–162.

[12] R. Hariharan, B. Hore, C. Li, and S. Mehrotra, "Processing

spatial- keyword (SK) queries in geographic information retrieval (GIR) systems," in *Proc. 19th Int. Conf. Sci. Statistical Database Manage.*, 2007, p. 16.

[13] S. Vaid, C. B. Jones, H. Joho, and M. Sanderson, "Spatio- textual indexing for geographical search on the web," in *Proc. 9th Int. Conf. Adv. Spatial Temporal Databases*, 2005, pp.218–235.

[14] A. Khodaei, C. Shahabi, and C. Li, "Hybrid indexing and seamless ranking of spatial and textual features of web documents," in *Proc. 21st Int. Conf. Database Expert Syst. Appl.*, 2010, pp.450–466.

[15] A. Guttman, "R-trees: A dynamic index structure for spatial searching," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 1984, pp. 47–57.

[16] I. De Felipe, V. Hristidis, and N. Rishe, "Keyword search on spatial databases," in *Proc. IEEE 24th Int. Conf. Data Eng.*, 2008, pp. 656–665.

[17] G. Cong, C. S. Jensen, and D. Wu, "Efficient retrieval of the top-k most relevant spatial web objects," *Proc. VLDB Endowment*, vol. 2, pp. 337–348, 2009.

[18] B. Martins, M. J. Silva, and L. Andrade, "Indexing and ranking in Geo-IR systems," in *Proc. Workshop Geographic Inf.*, 2005, pp. 31–34.

[19] Z. Li, H. Xu, Y. Lu, and A. Qian, "Aggregate nearest keyword search in spatial databases," in *Proc. 12th Int. Asia- Pacific Web Conf.*, 2010, pp. 15–21.

[20] M. L. Yiu, X. Dai, N. Mamoulis, and M. Vaitis, "Top-k spatial pref- erence queries," in *Proc. IEEE 23rd Int. Conf. Data Eng.*, 2007, pp. 1076–1085.

[21] T. Xia, D. Zhang, E. Kanoulas, and Y. Du, "On computing top-*t* most influential spatial sites," in *Proc. 31st Int. Conf. Very Large Databases*, 2005, pp. 946–957.

[22] Y. Du, D. Zhang, and T. Xia, "The optimal-location query," in *Proc. 9th Int. Conf. Adv. Spatial Temporal Databases*, 2005, pp. 163– 180.

[23] D. Zhang, Y. Du, T. Xia, and Y. Tao, "Progressive computation of the min-dist optimal-location query," in *Proc. 32nd Int. Conf. Very Large Databases*, 2006, pp. 643–654.

[24] N. Beckmann, H.-P. Kriegel, R. Schneider, and B. Seeger, "The R*- tree: An efficient and robust access method for points and rectangles," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 1990, pp. 322–331.

[25] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases," in *Proc. 20th Int. Conf. Very Large Databases*, 1994, pp. 487–499.

[26] P. Ciaccia, M. Patella, and P. Zezula, "M-tree: An efficient access method for similarity search in metric spaces," in *Proc. 23rd Int. Conf. Very Large Databases*, 1997, pp.426–435.

[27] R. Weber, H.-J. Schek, and S. Blott, "A quantitative analysis and perfomance study for similarity-search methods in high-dimensional spaces," in *Proc. 24th Int. Conf. Very Large Databases*, 1998, pp. 194–205.

[28] W. Johnson and J. Lindenstrauss, "Extensions of Lipschitz mappings into a Hilbert Space," *Contemporary Math.*, vol. 26, pp. 189– 206, 1984.

[29] J. M. Kleinberg, "Two algorithms for nearest-neighbor search in high dimensions," in *Proc. 29th Annu. ACM Symp. Theory Comput.*, 1997, pp. 599–608.

[30] A. Gionis, P. Indyk, and R. Motwani, "Similarity search in high dimensions via hashing," in *Proc. 25th Int. Conf. Very Large Data- bases*, 1999, pp. 518–529.

[31] V. Singh and A. K. Singh, "SIMP: Accurate and efficient near neighbor search in high dimensional spaces," in *Proc. 15th Int. Conf. Extending Database Technol.*, 2012, pp. 492–503.

[32] Y. Tao, K. Yi, C. Sheng, and P. Kalnis, "Quality and efficiency in high dimensional nearest neighbor search," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2009, pp.563— 576.

[33] H.-H. Park, G.-H. Cha, and C.-W. Chung,"Multi-way spatial joins using r-trees: Methodology and performance evaluation," in *Proc. 6th Int. Symp. Adv. Spatial Databases*, 1999, pp.229–250.

[34] D. Papadias, N. Mamoulis, and Y. Theodoridis, "Processing and optimization of multiway spatial joins using r-trees," in *Proc. 18th ACM SIGMOD-SIGACT-SIGART Symp. Principles Database Syst.*, 1999, pp. 44–55.

[35] T. Ibaraki and T. Kameda, "On the optimal nesting order for computing N-relational joins," *ACM Trans. Database Syst.*, vol. 9, pp. 482–502, 1984.

[36] D. Comer, "The ubiquitous b-tree," *ACM Comput. Surveys*, vol.11, no. 2, pp. 121–137, 1979.

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

An Efficient Method for High Quality and Cohesive Topical Phrase Mining

Anil
M.Tech Scholar,
CSE Department,
Malla Reddy College of Engineering
anil.suryapet@gmail.com

ABSTRACT

A phrase is a natural, meaningful, and essential semantic unit. In topic modeling, visualizing phrases for individual topics isan effective way to explore and understand unstructured text corpora. Usually, the process of topical phrase mining is twofold: phrase mining and topic modeling. For phrase mining, existing approaches often suffer from order sensitive and inappropriate segmentation problems, which make them often extract inferior quality phrases. For topic modeling, traditional topic models do not fully consider the constraints induced by phrases, which may weaken the cohesion. Moreover, existing approaches often suffer from losing domain terminologies since they neglect the impact of domain-level topical distribution. In this paper, we propose an efficient method for high quality and cohesive topical phrase mining. A high quality phrase should satisfy frequency, phraseness, completeness, and appropriateness criteria. In our framework, we integrate quality guaranteed phrase mining method, a novel topic model incorporating the constraint of phrases, and a novel document clustering method into an iterative framework to improve both phrase quality and topical cohesion. We also describe efficient algorithmic designs to execute these methods efficiently

INTRODUCTION

TOPICAL phrase mining refers to automatically extracting phrases which grouped by individual themes from given text corpora. It is of high value to enhance the power and efficiency to facilitate human to explore and understand a large amount of unstructured text data. One example is that if researchers could find phrases among a research field appearing with high frequencies in related proceedings in

different years, they will be able to have an insight into the academic trend of that research field. Topical phrase mining is not only an important step in established fields of

information retrieval and text analytics, but also is critical in various tasks in emerging applications, including topic detection and tracking [1], social event discovery [2], news recommendation system, and document summarization [3]. Usually, the process of topical phrase mining is twofold: phrase mining and topic modeling. These two stages not only directly affect the quality of discovered phrases and the cohesion of topics, but also, they may interact and indirectly impact each other's outcomes, e.g., low quality phrases (incomplete or meaningless) may cause misleading topical assignment in topic modeling. However, from phrase quality and topical cohesion perspectives, the outcomes of existing approaches remain to be improved. From phrase quality perspective, existing phrase mining methods [4–11] often produce low quality phrases. A high quality phrase should satisfy frequency, phrasegrness, completeness, and appropriateness criteria. Phrase mining is originated from the natural language processing (NLP) community, which utilizes predefined linguistic rules that rely on part-of-speech (POS) tagging or parsing trees [4, 5] to generate phrases. Such NLP based methods are commonly language-dependent and need texts to comply with grammar-rules, so it is not easy for them to be migrated to other languages and not suitable for analyzing some newly emerging and grammar-free text data, such as twitters, academic papers and query logs. In the hope to overcome the disadvantages of NLP based methods, there are many data-driven approaches that have been proposed in this area. Data-driven methods primarily view phrase mining as a frequent pattern mining problem [6, 7]. A phrase is extracted if it is constituted by the longest word sequence whose frequency is larger than a given threshold. Inevitably, extracting word sequence according to frequency is prone to produce many false phrases. Recently, researchers have sought for a kind of general, yet powerful phrase mining method. A variety of statistic-based methods [8–10] have been proposed to improve phrases quality by ranking

ISBN 978-93-88808-61-3

candidate phrases. A more recent work [11] considers integrating phrasal segmentation with phrase quality estimation to estimate rectified phrase frequency to further improve phrase quality.

However, due to suffering from order sensitive and inappropriate segmentation, the outcome of existing methods is still inadequate. Below we use Table 1 to show the deficiencies of the existing methods by using significance scores Sig score extracted from a corpus, 5Conf. 1 We compared two phrases using different processing orders based on 5Conf. Data in Table 1 is derived from the result of an existing method [9] which heuristically merges words under t-test score (i.e., a statistical hypothesis test to measure whether its actual occurrence significantly different from expected occurrence). The expected occurrence of phrase Pr = w1 _ w2 is calculated by f(w1)_f(w2) N , where f(w1) and f(w2) are word frequencies of w1 and w2 in the corpus, respectively, and N is the total number of words in the corpus. The method [9] allows users to specify a threshold of a significance score Sig score(Pr) of a phrase Pr, which is the statistical significance of taking a group of words as a phrase. It is measured by comparing the actual frequency with the expected occurrence. A larger value of Sig score(Pr) indicates the word sequence Pr has higher possibility to be a whole unit (phrase) than other sequences, and vice versa.

(1) Order sensitive. Assume Gaussian Mixture Model is a high quality phrase since it is complete in semantic. By choosing the merge order1 2 :3 , as shown in Table 1, existing approaches heuristically merge Gaussian and Mixture firstly, since the order shows a higher t-test score 6391:62 to achieve a local optimum comparing with the score 23:96 by using the order 2 3 :1 . However, if the threshold Sig score = 16, the complete phrase Gaussian Mixture Model failed to be extracted by using the order 1 2 :3 since the final core 15:75 is less than the given threshold 16 (we use symbol _ to

ISBN 978-93-88808-61-3

denote the score of the whole phrase under the given merge order). Instead, the merge order 2 3 :1 could have this phrase extracted. For the second phrase Peer to Peer Data, by using the same corpus, we got the same conclusion. Consequently, the completeness of extracted phrases highly depends on the merging order of the merging heuristics. The incompleteness brought by traditional approaches will cause incomplete semantics and may produce very general phrases. For instance, phrase Mixture Model has many explanations, such as Gaussian Mixture Model, Finite Mixture Model, or Interactive Mixture Model, whereas by phrase Gaussian Mixture Model, one explicitly refers to the very probabilistic model.

(2) Inappropriate segmentation. For the word sequence Gaussian Mixture Model Selection, it contains two quality phrases Gaussian Mixture Model and Model Selection since they both have high statistic scores. However, these two quality phrases are overlapping in the sequence. In the scenario of text chunking, the word

model can only belong to one of these two phrases, i.e., s1 = Gaussian Mixture Model | Selection or s2 = Gaussian Mixture | Model Selection. Existing approaches which only consider intra-coocurrence (e.g., phrase frequency and phrase length) prefer to choose sequence s2 , since both Gaussion Mixture and Model Selection have high frequencies. However, Gaussian Mixture Model should be the right choice for it is a whole function unit as an adjective, while Gaussian Mixture is obviously an incomplete phase.

From topical cohesion perspective, traditional topic models, such as LDA, assume words are generated independently from each other, i.e. “bag-of-words” assumption. Under this assumption, a phrase is regarded as an independent “word”, which may lead to the loss of its specific meaning, and as a result, the impact of phrases is ignored. To address the topic assignment problem associated with phrase, some existing methods

ISBN 978-93-88808-61-3

such as PhraseLDA [9] uses an undirected clique to model the stronger correlation of words in the same phrase on top of the "bag-of-phrases" assumption. To be specific, words in the same phrase form a clique, and PhraseLDA imposes the same latent topic on the words in the same clique. However, it is not enough to consider only the correlation of a phrase and its words. A phrase as a whole may carry lexical meaning that is beyond the sum of its individual words. For example, the phrase max pooling has a meaning beyong the word "max" or "pooling". Thus, it would be inappropriate to enforce words in the same phrase to inherit the same topic like PhraseLDA does, since long noun phrases sometimes do have components indicative of different topics [12].

Moreover, existing approaches neglect a fact that some phrases are only valid in certain domains. Usually, the texts within a corpus often come from more than one domain, and each domain may contain its own terminologies. These domain-specific terminologies may only appear frequently within certain domains but not in others, making them less possible to be extracted in the entire corpus where their occurrence frequency is diluted by the other domains, asTable 2 demonstrates.

In Table 2, the phrases support vector machine, eigen vector, bit vector, and social networks are estimated to belong to machine learning (ML) , math (MA), database (DB), and data mining (DM) domains, respectively. Even though some phrases (e.g., support vector machine and social networks) can achieve a high enough significance in the entire corpus, while others such as bit vector and eigen vector cannot. Consequently, it is hard for them to be mined as phrases in the entire corpus, albeit actually they both are common terminologies in their own domains.

Besides effectiveness, efficiency is also very important to topical phrase mining, especially for the applications that need timely analysis, such as topic-tracking [1], social event

discovery [2], and news recommendation system. Take Twitter as an example, the volume of tweets grew at increasingly high rates from its launch in 2006 to 2010, approaching around 1; 000% gain in yearly volume2. Currently, over 350; 000 tweets are generated on Twitter per minute. Unfortunately, most existing approaches [11–14] often suffer from low efficiency as they cannot support such high throughput tasks.

In order to effectively and efficiently mine topical phrases and improve phrase quality and topical cohesion, we propose a Cohesive and Quality Topical Phrase Mining (CQMine) framework, which automatically clusters documents with a more sensible topic model, and improves the quality of phrases by adopting more accurate and rigorous mining approaches. Moreover, our quality phrase mining approach can be solely used to mine phrases. The main contributions of this paper are as follows:

We propose effective and efficient quality phrase mining approaches. By eliminating order sensitive and avoiding inappropriate segmentation, our approaches could guarantee the quality of extracted phrases. Moreover, we also design effective algorithms to accelerate the processing.

We propose a novel topic model to address topic assignment problem associated with idiomatic phrases to improve the cohesion of topical phrases. Considering the fact that some phrases are only valid in certain domains, we propose an iterative framework to facilitate more accurate domain terminologies finding. _ Experimental evaluation and case study demonstrate that our method is of high interpretability and efficiency compared with the state-of-the-art methods.

Existing System

Topical phrase mining is not only an important step in established fields of information retrieval and text analytics, but also is critical in various tasks in emerging applications,

ISBN 978-93-88808-61-3

including topic detection and tracking , social event discovery , news recommendation system, and document summarization .the process of topical phrase mining is twofold: phrase mining and topic modeling. These two stages notonly directly affect the quality of discovered phrases and the cohesion of topics, but also, they may interact andindirectly impact each other's outcomes, e.g., low quality phrases (incomplete or meaningless) may cause misleading topical assignment in topic modeling. However, from phrase quality and topical cohesion perspectives, the outcomes of existing approaches remain to be improved.

NLP based methods are commonly language-dependent and need texts to comply with grammar-rules, so it is not easy for them to be migrated to other languages and not suitable for analyzing some newly emerging and grammar-free text data, such as twitters, academic papers and query logs. In the hope to overcome the disadvantages of NLP based methods, there are many data-driven approaches that have been proposed in this area. A variety of statistic-based methods have been proposed to improve phrases quality by ranking candidate phrases.

Proposed System

Considering the fact that some phrases are only valid in certain domains, we propose an iterative framework to facilitate more accurate domain terminologies finding. Experimental evaluation and case study demonstrate that our method is of high interpretability and efficiency compared with the state-of-the-art methods.

Future Work

Different with the existing model which only considers intra-co occurrence of phrases and regards the generation of segmentations as an independent process. Our methods comprehensively consider both the intra-co occurrence of phrases and the isolation of partition position. From a technical perspective, the isolation of "current" split position depends on the "future" generated split position. Thus, we need to check every possible new split positions to determine the isolation of current split position, which makes the computation of optimal segmentations very time consuming. To address this issue, we adopt a dynamic programming strategy, which is based on an observation that if bi+1 and the previous partition position bi is the optimal position.

News Publisher

News publisher provides the news articles on daily basis, breaking news; live news etc. news data are stored in

ISBN 978-93-88808-61-3

database. Offering the services to the end users. News Recommendation system publish the news articles based on categories. News Publisher search the news topics randomly whether the articles are displaying related to category. Users Registered in news portal to view the news articles, once read the article can also to comment the article and shared to others

Effectiveness Analysis of quality phrase

Examined the effectiveness of our quality phrase mining stage by measuring the phrase quality in two metrics: (1) Wiki-phrases benchmark and (2) Expert Evaluation. **Wiki-Phrases:** Wiki-phrases is a collection of popular mentions of entities by crawling intra-Wiki citations within Wiki content. Wiki phrases benchmark provides a good coverage of commonly used phrases which could avoid the variance caused by different human raters. In this evaluation,we regarded Wiki phrases as ground truth phrases. That is to belongs to/not belongs to Wiki phrases. To compute precision, only the Wiki phrases are considered to be positive. For recall, we firstly mergedall the phrases returned by all methods including ours, and then we obtained the intersection between the Wiki phrases and the merged phrases as the evaluation set.

Quality Phrase Mining

In the CQMine framework the quality phrase mining stage contains three steps:
Firstly, a PhraseTrie is built to count all possible phrases' frequencies. Then, a complete phrase mining algorithm is applied to mine complete phrases, which will be under the guidance of a statistics-based measurement to satisfy phraseness criterion. During phrase mining, the mined phrases are stored inPhraseTrie to avoid recomputing duplicate phrases. Finally, to guarantee the appropriateness requirement, for each document, CQMine needs to check if it contains overlapping phrases, if so, we will partition them into non-overlapping phrases by utilizing an effective and efficient overlaping phrases segmentation algorithm. After quality phrase mining, a document is transformed from a multiset of words (bag-of-words) into a multiset of phrases (bag-of-phrases) which will be taken as the input of topic modeling.

Topical phrase mining

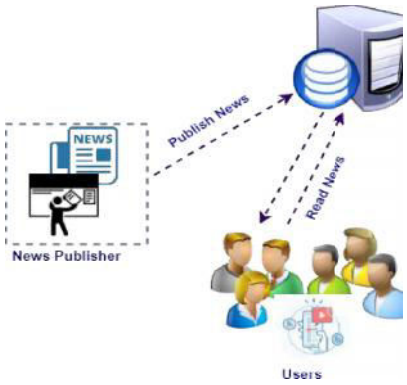
Significant progresses have been made on the topical phrase mining and they can be broadly classified into three types:

ISBN 978-93-88808-61-3

(1) Joint learning phrases and their topic assignment,
(2) Mining phrases posterior to topic inferring,
(3) Mining phrases prior to topic inferring.

Word sequence segmentation (or phrasal segmentation) is another strategy for phrase mining. Formally, phrasal segmentation aims at partitioning a word sequence into a set of disjoint subsequences, each indicating a phrase. It only considers intra co occurrence of phrases such as phrase length and words, while ignores the inter-isolation between phrases. The second strategy utilizes a post-processing step to generate phrases after inferred by the LDA model. Recursively merges consecutive words with the same latent topic by a distribution-free permutation test on arbitrary length back-off model until all significant Consecutive words have been merged. it performs phrase mining and topic inferring simultaneously by incorporating successive word sequence assumption into the generative model. Wallach proposed a bigram topic model based on a hierarchical Dirichlet allocation model. Bigram model is a probabilistic generative model that conditions on the previous word and topic when drawing the next word.

Architecture



Algorithm

The completeness of extracted phrases highly depends on the merge order. In order to obtain the complete phrases, we need to enumerate every possible merge order. Obviously, a straight-forward algorithm of finding the complete phrases in document d is: enumerating all the subsequences of this document first, then verify whether each one is a complete phrase.The algorithm QBA (q-Chunk Based Approach) firstly generates boundaries It then computes the local solution of each chunk using DPBA denote the left boundary of current chunk. For each boundary algorithm QBA checks whether satisfies merge condition.

The main processing steps of QBA are as follows:
(1) Partitioning the sequence into a series of q-length chunks;
(2) Performing top-down search on each chunk to get local solutions

(3)Checking whether two adjacent chunks need to be merged.

If they do not need to be merged, it means no phrase could cross the boundary between the two chunks. Otherwise the two chunks are merged into a new chunk and QBA will find new solutions on the new chunks.

CONCLUSIONS

We presented an efficient method for cohesion and quality topical phrase mining. In phrase mining stage, we focus on quality phrase mining problem, and propose two efficient quality phrase mining algorithms. In practice, the time cost of our best exact algorithm is competitive to greedy algorithm. In topic modeling stage, we propose a novel topic model to incorporate the constraint that is induced by phrases; moreover, it can well address the collocation phrase issue. Finally, considering the fact that some phrases are only valid in certain domains, we cluster documents under the condition that they share similar topic distribution and iteratively perform cluster updating and topical inferring to further improve the cohesion of topical phrases. The empirical verification demonstrated our framework has high interpretability and efficiency.

REFERENCES

1. J. Leskovec, L. Backstrom, J. Kleinberg, "Meme-tracking and the dynamics of the news cycle", *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, pp. 497-506, 2009.

2. M. Li, J. Wang, W. Tong et al., "EKNOT: Event knowledge from news and opinions in twitter", *Proc. 30th AAAI Conf. Artif. Intell.*, pp. 4367-4368, 2016.

3. Z. He, C. Chen, J. Bu et al., "Document summarization based on data reconstruction", *Proc. AAAI Conf. Artif. Intell.*, pp. 620-626, 2012.

4. S. P. Abney, "Parsing by chunks" in Principle-Based Parsing, The Netherlands:Kluwer Academic Publishers, pp. 257-278, 1991.

5. H. Clahsen, C. Felser, "Grammatical processing in language learners", *Applied Psycholinguistics*, vol. 27, no. 27, pp. 3-41, 2006.

6. M. Danilevsky, C. Wang, N. Desai et al., "Automatic construction and ranking of topical keyphrases on collections of short documents", *Proc. Int. Conf. Data Mining*, pp. 398-406, 2014.

7. A. Simitsis, A. Baid, Y. Sismanis et al., "Multidimensional content exploration", *Proc. VLDB Endowment*, vol. 1, no. 1, pp. 660-671, 2008.

8. A. Parameswaran, H. Garcia-Molina, A. Rajaraman, "Towards the web of concepts: Extracting concepts from large datasets", *Proc. VLDB Endowment*, vol. 3, no. 1–2, pp. 566-577, 2010.

9. A. El-Kishky, Y. Song, C. Wang et al., "Scalable topical phrase mining from text corpora", *VLDB Endowment*, vol. 8, no. 3, pp. 305-316, 2014.

10. P. Deane, "A nonparametric method for extraction of candidate phrasal terms", *Proc. 43rd Annu. Meeting Assoc. Comput. Linguistics*, pp. 605-613, 2005.

11. J. Liu, J. Shang, C. Wang et al., "Mining quality phrases from massive text corpora", *Proc. ACM SIGMOD Int. Conf. Manage. Data*, pp. 1729-1744, 2015.

12. X. Wang, A. McCallum, X. Wei, "Topical N-grams: Phrase and topic discovery with an application to information retrieval", *Proc. 7th IEEE Int. Conf. Data Mining*, pp. 697-702, 2007.

13. C. Wang, M. Danilevsky, N. Desai et al., "A phrase mining framework for recursive construction of a topical hierarchy", *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, pp. 437-445, 2013.

14. R. V. Lindsey, W. P. Headden, M. J. Stipicevic, "A phrase discovering topic model using hierarchical pitman-yor processes", *Proc. Joint Conf. Empirical Methods Natural Language Processing Comput. Natural Language Learn.*, pp. 214-222, 2012.

15. E. Pitler, S. Bergsma, D. Lin et al., "Using web-scale N-grams to improve base NP parsing performance", *Proc. 23rd Int. Conf. Comput. Linguistics*, pp. 886-894, 2010.

16. M. F. Porter, Snowball: A Language for Stemming Algorithms, Palo Alto, CA, USA:Open Source Initiative Osi, 2001.

17. M. F. Porter, "An algorithm for suffix stripping", *Program*, vol. 14, no. 3, pp. 130-137, 1980.

ISBN 978-93-88808-61-3

18. K. Pearson, "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling", *Philosophical Magazine*, vol. 50, no. 302, pp. 157-175, 1900.

19. T. L. Griffiths, M. Steyvers, "Finding scientific topics", *Proc. Na. Academy Sci. United States America*, vol. 101, no. 1, pp. 5228-5235, 2004.

20. S. Geman, D. Geman, "Stochastic relaxation gibbs distributions and the Bayesian restoration of images", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721-741, Nov. 1984.

21. B. Li, B. Wang, R. Zhou et al., "CITPM: A cluster-based iterative topical phrase mining framework", *Proc. Int. Conf. Database Syst. Advanced Appl.*, pp. 197-213, 2016.

22. S. Kullback, R. A. Leibler, "On information and sufficiency", *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79-86, 1951.

23. A. Rodriguez, A. Laio, "Clustering by fast search and find of density peaks", *Sci.*, vol. 344, no. 6191, pp. 1492-1496, 2014.

24. K. Frantzi, S. Ananiadou, H. Mima, "Automatic recognition of multi-word terms: The C-value/NC-value method", *Int. J. Digital Libraries*, vol. 3, no. 2, pp. 115-130, 2000.

25. I. H. Witten, G. W. Paynter, E. Frank et al., "KEA: Practical automatic keyphrase extraction", *Proc. ACM Conf. Digital Libraries*, pp. 254-255, 1999.

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

Nearest Keyword Set Search Queries on Multi-Dimensional Datasets

Mr.Rajesh,
M.Tech.
Scholar,
Department of CSE,
Malla Reddy College of Engineering
rajcherukumalli@gmail.com

Abstract – **Data mining is an important aspect in refining the data, this makes the data simpler and easy to use and it is a widely used technique to extract the useful information from huge chunks of data. The application developed is a user friendly application which helps to extract the informative keywords from text files and associates keywords with the file and the searching and retrieval of file is made easier with nearest keywords in a multidimensional data set of files. The integration of data mining concepts along with the cloud storage security techniques lead to the development of an efficient search application to retrieve files from the repository using easily remembered keywords.**

Keywords: Data security, keyword extraction, multidimensional data set.

I. INTRODUCTION

Present world is a technology centric world where each and every small works carried out in day to day life is dependent on the technology. Technology is rapidly growing such that it has occupied major portion in human routine. In this technological era the dependency of people over the information is growing hand in hand with the technology. Each and every activity is dependent on accessing the information and processing with it to get the required result. Starting from a small home to large organisation storing and retrieving of information has become an inseparable part of one's routine. Researches' have stated that on a whole the collection of data includes 10% of structured data and the remaining 90% are unstructured data. Information retrieval from structured data set is simpler compared to information retrieval from unstructured datasets. In order to discover the hidden patterns in the unstructured dataset data mining techniques are used.

Data mining is the process in which the patterns are extracted from the data set and the discovered patterns are analysed and used in studies. Machine learning, statistical analysis are some of the domains which uses the data mining techniques. In this application the data mining is applied over the text documents. The text documents with varied contents without any specific patterns to extract are being considered. A multidimensional data is the type of data in which there is heterogeneous data type objects that are grouped under certain attributes. But when the text document with no specified patterns are considered the pattern recognition becomes impossible, in such cases the Keywords extracted from every document is one dimension of the document that describes the feature of the document.

II. MOTIVATION

Over time the collection of documents increases and remembering all file names is impossible for normal human brains and retrieval of file is impossible if file name is unknown, but one can actually be aware of the searching content that is present in a file. In this application the searching based on the keywords document with the associated keywords are easier to search as data consumer can easily interpret about the content of the document rather than the file name. An application to search the documents along with the secured storage for documents is very necessary in every organisation and its implementation is shown in section four and the literature study is shown in section three, design in section four and conclusion in section six.

III. LITERATURE SURVEY

First technique is and it mainly on computing exact nearest and farthest neighbour which is a challenging task, especially in the case of high-dimensional data. Many techniques are used to solve the nearest neighbour problem but not much importance is on farthest neighbour problem. By the calculation of the farthest neighbour a clear idea is obtained for the elimination of unrelated objects to the query there by it helps in giving the result more accurately.[1]

Multidimensional text cube analysis is another technique which is used to analyse the textual documents and the analysis is done by applying the data mining technique over the documents in order to extract the hidden patterns out of it.[2]

Collective spatial keyword technique is another technique which is used to derive the result such that more than one object is required to satisfy the user's queryin such cases one node is being considered as owner and other two nodes as sub objects that matches closely with query keyword. This is helpful in developing navigational search query applications where more than one object is necessary to satisfy the user's need.[3]

Cloud computing being one of the blooming technologies provides various services one such facility is the storage at minimum cost which makes most of the

ISBN 978-93-88808-61-3

organizations to store their data inside the cloud but there are chances of cloud being vulnerable to attacks. There is a technique called n-keyword search where the n is total number of distinct keywords present in all the documents and scalar product is performed over the co-ordinate keywords and homomorphic key exchange is done between the participating entities such that transmission is done securely. [4]

The literature survey of these papers helped in developing an idea to implement a search application over multidimensional data set.

IV. SYSTEM DESIGN

Any project before being implemented must be designed such that it gives a complete view of the entire project. The workflow is the main part of project design which is step wise explanation to the flow of project. The flow diagram for the project is shown below in Figure 1.

Figure 1 shows the work flow of admin module. There are six operations carried out at the time of uploading a file. The specified file is selected and file is being sent in to stop words removal and stemming process informative keywords are fetched and keyword filtration is done. The file is uploaded along with its associated keywords. The workflow of user is shown in figure 2 once user registers and gets the user ID and password and decryption key to the mail box from admin, then user can search the file by decrypting the key given to him and search file with the keywords and download if needed.

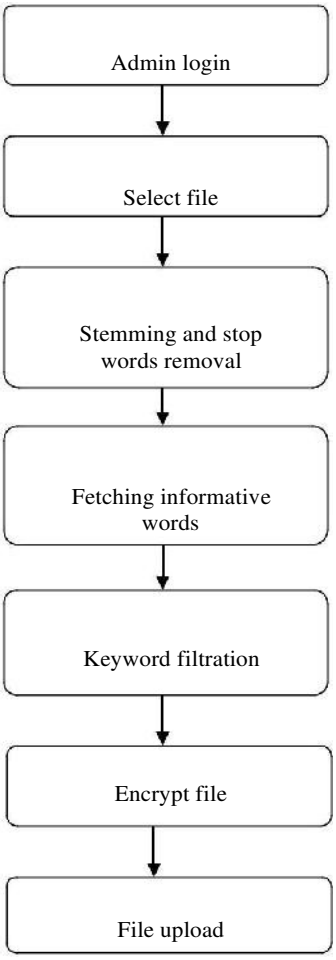


Figure 1 Work flow of admin

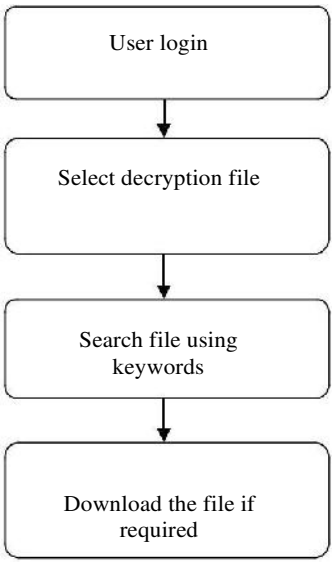


Figure 2 Work flow of user

Algorithm for upload process

Step 1: Start
Step 2: Read File (F) and access permission based on category(C).

Step 3: Remove unnecessary words and special characters. Step 4 : Shortlist the Keywords. Step5:UsingTermFrequency(TF) calculates weight age for Keywords.

Step 6: Let N be the number of category allowed to access. Step 7: For I=1 To N.
Step 8: Fetch the hash key of Ith category.
Step 9: Using hashing technique with fetched hash key,

generate keyword hash tags for all the keywords.

Step 10: Insert all keyword hash tag into index. Step 11: Repeat from Step 7 to Step 10 up to I=N. Step 12: Stop

Algorithm for search process

Step 1: Start
Step 2: Get the Keyword (K) from User (U).
Step 3: Find the User Category (UC).
Step 4: Fetch Hash key for UC.
Step 5: Generate Trapdoor using UC hash key.
Step 6: Search trapdoor on index array.
Step7: Filter all the matched index elements.
Step 8: Shortlist filter from filtered index.
Step 9: Using Inverse document Frequency Rank the files.
Step 10: Display the file list to user.
Step 11: Stop

ISBN 978-93-88808-61-3

The above are the algorithms for upload and search processes respectively. The upload process is carried out by the data provider and downloading happens by the data consumer.

V. PROPOSED MODEL

The system architecture of the search application developed is shown in figure 3. The two main modules are admin and the user. This is a multiuser environment based application. Once the user is authorised the user can get access to the files through the internet if the application is installed in the user's system. Admin is the data owner who maintains the user details and uploads the files in to the database. Once the registration is done the user gets a decryption sent to the mail id. If a user wants to retrieve the files then the decryption key sent to the user's mail is decrypted then the search box opens up and the user can search the file using the predicted keywords.

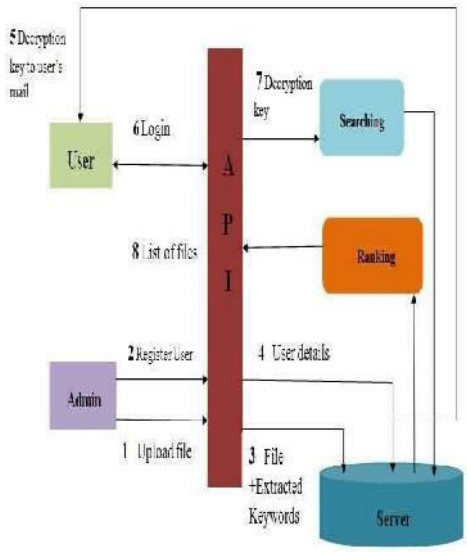
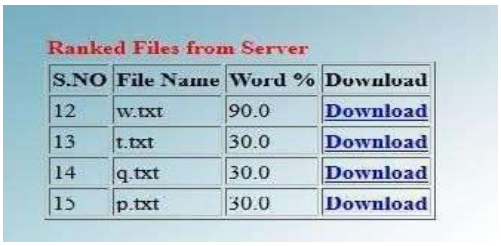


Figure 3: System architecture

Admin being the data owner maintains records and at the time of uploading files the admin is given the authority to select the grade such that the specific file being uploaded can be viewed only by the user who belongs to same grade. This facility is provided with a security point of view and also to reduce the searching process which happens only within the particular grade to which user belongs. The searching process need not consider all the files inserted in the database instead it is sufficient to carry out the searching only within the particular grade. Once the user logs in the user needs to decrypt the key which is provided by the admin at the time of the registration. Once the decryption of key is done then the search box is being opened to the user, where the user can type the keywords either multiple or single keyword. Once the matching of keywords is done the respected files are displayed such that the file with the highest rank will be at the top followed by files with least ranking.

VI. RESULT ANALYSIS

The result of the work carried out is to get the list of files based on the keywords weightage in the form of ranked list as shown in figure 4. The test cases for the work are shown in figure 5.



Ranked Files from Server

S.NO	File Name	Word %	Download
12	w.txt	90.0	Download
13	t.txt	30.0	Download
14	q.txt	30.0	Download
15	p.txt	30.0	Download

Figure 4 : List of files

Test cases	Input	Result
Keywords	Keywords matched with index keywords	PASS
	Keywords spelt wrong or not matched	FAIL
Number of keywords	Keywords less than five	PASS
	Keywords more than five	FAIL
Type of documents	Text file	PASS
	Pdf, word document, excel, image file.	FAIL

Figure 5 : Table of test cases

The analysis is done based on the uploading and downloading time. The uploading time take for any file size is always greater than the downloading time. Uploading takes more as the stemming and stop words removal are happening at the time of uploading. And the downloading time is lesser than the uploading time this because there is no much steps involved in downloading a file hence file search is efficient using this application.

The time recorded for various sizes is shown in figure 6 and the bar chart in figure 7 clearly depicts that for any size of files the uploading time is always greater than downloading time.

VII. CONCLUSION

The application developed is useful in any organization where the huge amount text data is stored in the form of files, so late at the time of retrieval the work of user becomes easier by searching the file with the keywords. The application can further be developed to take inputs such as images, audio, video and documents other than text files and also the application can be developed in to an android app which reduces the burden of the user to a greater level.

VIII. RRFERENCE

[1] Yifan Hao, Huiping Cao, Yan Qi, Chuan Hu, Sukumar Brahma, Jingyu HanNew Mexico State University, Las Cruces, "Efficient Keyword Search on Graphs using MapReduce" 2015 IEEE International Conference on Big Data (Big Data)

[2] Cheng Long Raymond Chi-Wing Wong Ke Wang Ada Wai-Chee Fu The Hong Kong University of Science and Technology Simon Fraser University The Chinese University of Hong Kong "Collective Spatial

[3] Shafin Rahman Department of Electrical & Computer Engineering North South University Dhaka, Bangladesh Mrigank Rochan Department of Computer Science University of Manitoba Winnipeg, Canada "Fast Farthest Neighbor Search Algorithm for Very High Dimensional Data" 19th International Conference on Computer and Information Technology, December 18-20, 2016, North South University, Dhaka, Bangladesh

[4] Fangbo Tao, Kin Hou Lei, Jiawei Han, ChengXiang Zhai, "EventCube: Multi-Dimensional Search and Mining of Structured and Text Data" KDD'13, August 11–14, 2013, Chicago, Illinois,USA.
Copyright 2013 ACM 978-1-4503-2174-7/13/08

[5] Gandeevan Raghuraman Pooja Nilangekar Pallavi Vijay Kavya Premkumar Saswati Mukherjee "Cloud based Privacy Preserving EfficientDocument Storage and RetrievalFramework"1234Department of Computer Science,Anna University,Chennai 5Department of Information Technology, Anna University Chennai 978-1-4673-8200-7/15/ ©2015 IEEE

Proceedings on International Conference on Emerging Technologies in Computer Science(ICETCS-2019)
ISBN 978-93-88808-61-3

Analyzing And Detecting Money-Laundering Accounts in online social networks

Thati
Sindura
M.Tech
Scholar
Malla Reddy College of Engineering Hyderabad

Abstract

Virtual currency in OSNs plays an increasing- ly important role in supporting various financial activities such as currency exchange, online shop- ping, and paid games. Users usually purchase virtual currency using real currency. This fact motivates attackers to instrument an army of accounts to collect virtual currency unethically or illegally with no or very low cost and then launder the collected virtual money for massive profit. Such attacks not only introduce significant financial loss of victim users, but also harm the viability of the ecosystem. It is therefore of central importance to detect malicious OSN accounts that engage in laundering virtual currency. To this end, we extensively study the behavior of both malicious and benign accounts based on operation data collected from Tencent QQ, one of the largest OSNs in the world. Then, we devise multi-faceted features that characterize accounts from three aspects: account viability, transaction sequences, and spatial correlation among accounts. Finally, we propose a detection method by integrating these features using a statistical classifier, which can achieve a high detection rate of 94.2 percent at a very low false positive rate of 0.97 percent.

Introduction

Online social networks (OSNs) have started to leverage virtual currency as an effective means to glue financial activities across various platforms such as online shopping, paid online games, and paid online reading. Examples of virtual currency in such OSNs include but are not limited to Ten- cent Q Coin, Facebook Credits[1], and Amazon Coin. Usually, users purchase virtual money using real currency at a regulated rate; one user can also transfer it to another user via various meth- ods such as recharging their account and send- ing gifts [1]. These facts enable attackers to gain potentially massive profits through the following steps. First, an attacker can collect virtual currency with zero or low cost. For example, they can com- promise and subsequently control a legitimate account or register a huge number of accounts to win gifts (in the form of virtual currency) in online promotion activities. Next, they can instru- ment accounts under their control to transfer vir- tual currency to other accounts in return for real currency, with rates that are usually much lower compared to the regulated rate. Attackers usually post advertisements in popular e-commerce web- sites [2] to attract potential buyers. We call OSN accounts that are used by attackers for the collec- tion and transfer of virtual currency *MONEY-LAUN-dering ACCOUNTS*. Money-laundering accounts have caused a tremendous financial loss for compro- mised accounts, fundamentally undermined the effectiveness of online promotion activities, and possibly introduced potential conflicts against cur- rency regulations.

Detecting money-laundering accounts in OSNs therefore becomes of essential importance, which, however, is faced with new, significant challenges. First, committing money- laundering activities does not require the use of traditional malicious content such as spam, malicious URLs, or malicious executables. Although spamming might be used by attackers for advertising, neither methods nor the accounts used for spamming are necessarily

associated with the money-launder- ing accounts. Second, money-laundering activi- ties do not rely on social behavior and structures (e.g., “following” or “friend” relationship in popu- lar social networks) to operate. These challenges make existing methods immediately ineffective, since they focus on detecting OSN-based spam- ming, phishing, and scamming attacks, whose proper operation necessitates malicious content [3, 4], social structures [5], or social behaviors [6]. Detecting money laundering activities in tra- ditional financial transactions has attracted signif- icant research efforts [7]. For example, Dreewski *et al.* [8] designed a system to detect money laun- dering activities from billings and bank account transactions. Paula *et al.* [9] used the AutoEncod- er to classify exporters and detect money launder- ing activities in exports of goods and products in Brazil. Colladon *et al.*
[10] presented predictive models to quantify risk factors of clients involved in the factoring business and proposed a visual analysis method to detect the potential clusters of criminals and prevent money laundering. Dif- ferent from traditional money laundering detec- tion problems in bank-related activities, account behaviors of laundering virtual currency in OSNs involve bank-related financial activities, online social networks, and virtual recharging and expen- diture activities.

The goal of our work is to design an effective method capable of detecting money- laundering accounts. As a means toward this end, we per- form an extensive study of behaviors of mon- ey-laundering accounts based on data collected from Tencent QQ, one of the largest OSNs in the world with a giant body of reportedly 861 million active users. We have devised multi-facet- ed features that characterize accounts from three aspects: account viability, transaction sequences, and spatial correlation among accounts. Experi- mental results have demonstrated that our meth- od can achieve a high detection rate of
94.2 percent with a very low false positive rate of 0.97
percent. To the best of our knowledge, this work represents the first effort to analyze and detect money-laundering accounts in OSNs integrating virtual currency at this large scale.

Data Set

We have collected labeled data from Tencent QQ, a leading online social network in China, which offers a variety of services such as instant messaging, voice chat, online games, and online shopping. These services are glued together using Q coin, the virtual currency distributed and man- aged by Tencent QQ. Tencent QQ has a giant body of 861 million active accounts with a report- ed peak of 266 million simultaneously online users. Also, Tencent QQ is one of the leading OSNs that are actively involved in virtual currency based services in the world.

Our data set is composed of 114,891 mali- cious accounts and 381,523 benign accounts that are active during the first week of August in 2015. In order to label accounts used for money laundering, we follow advertisements of cheap virtual currency in major e-commerce websites and actually purchased virtual currency from sell- ers, where QQ accounts used by these sellers are labeled as money- laundering accounts. Since an attacker usually controls a large number of mali- cious accounts for money laundering, we label accounts as malicious if they login from the same IP address used by a confirmed money-laundering account within one day.

Although this labeling process offers us the ground truth, using it as a detection method is practically challenging. First, it requires a consider- able amount of investment to engage money-laun- dering accounts in malicious activities. Second, the IP addresses used to label launder accounts usually will be invalid after a few days, because attackers change the

ISBN 978-93-88808-61-3

login IP addresses frequent- ly. Therefore, this data labeling process, if used as a detection method, cannot guide OSNs to mitigate their financial loss proactively.For each account, we collect the following activity records. It is worth noting that all these records can be col- lected from social networks that integrate virtual currency.

- Login activities, which include the account ID, the login date, the login IP address, and the account level.
- Expenditure activities, which include the expenditure account ID, the expenditure date, the expenditure amount, the pur- chased service, the payment method, and the account ID to receive the service.
- Recharging activities, which include the recharging account ID, the recharging date, the recharging amount, and the payment method.

Behavior analysis and feature extraction

Figure 1 shows a typical process of virtual currency laundering. The first step is to collect virtual currency with zero or extremely low cost. For example, attackers can hack users' accounts (and thus control their virtual currency), exploit the system vulnerabilities, or participate in online promotion activities to win virtual currency for free or at significantly discounted rates [2]. Next, attack- ers attract potential buyerswith considerable

centage of benign accounts (less than 20 percent) experience the same activity level (i.e., being activefor less than 10 percent of total days).

Next, we study the source of virtual currency for benign and laundering accounts. A benign user usually recharges their account via wire trans- fer (often in the form of mobile payment) and occasionally receives gifts (from friends). Comparatively, money-laundering accounts almost exclusively rely on online promotions to directly collect virtual currency or gifts transferred from other accounts. We therefore introduce the following feature to characterize the currency collection behavior.

Feature 3: Percentage of Recharge from Mobile Payment: This feature represents the per- centage of virtual currency recharged through mobile payments (i.e., purchasing virtual currency using mobile online banks).

Figure 2c presents the distribution for this feature, where approximately 24 percent of benign users recharge their accounts via mobile payment, while the vast majorit y of malicious accounts do not use this channel.

As an increasing number of financial functions areintegrated into social networks, usersconduct a variety of activities such as shopping and gift- ing. While benign users prefer to engage financial activities with higher diversity, money- laundering accounts only focus on activities relevant to laundering. Therefore, we introduce the following five features to characterize such a difference.

FEATURE 4: The NUMber of Self-EXPENDITUREs: This feature represents the total number of expenditures that an account has committed to itself using virtual currency.

FEATURE 5: The NUMBER of EXPENDITURES Not from VIRTUAL CURRENCY: This feature characteriz- es the number of expenditures an account has committed by other methods instead of virtual currency.

Feature 6: Percentage of Expenditure from Banks: A user can associate their bank account with the OSN account. This bank account can be directly used for shopping and gifting in addition to virtual currency in the OSN account. This feature is defined as the percentage of expenditure from associated bank accounts.

Feature 7: The NUmber of AccoUnts that ever Receive Gifts from this ACCOUNT:

ISBN 978-93-88808-61-3

Malicious accounts need to frequently transfer the virtual currency as a gift to the buyer accounts, while a benign user tends to expend the virtual currency themself, and occasionally gives the virtual cur- rency as a gift to their friends. Thus, malicious accounts will have a much larger value of this fea- ture than benign users.

FeatUre 8: Percentage of the AmoUnt of ExpenditUres as Gifts: This feature represents the proportion of the amount of expenditures as gifts in all expenditures. After malicious accounts collect virtual currency from the online promo- tion activities and other vulnerabilities, they will transfer it to other accounts as gifts. We therefore introduce this feature to quantify the percentage of all giving out behavior.

Figures 2d–2h report the distributions for Features 4–8 respectively. Almost all the malicious accounts (more than 99 percent) neither committed for itself using virtual currency nor committed by other methods instead of virtual currency. Comparatively,

61 percent of benign accounts have committed for itself using virtual currency at least once, and 18 percent of benign accounts have committed by other methods instead of vir- tual currency at least once. The distributions for Features 6–8 are also distinguishable as shown in the figures. We omit the descriptions for brevity.

Sequential AI Features of Financial Activities

The sequences of financial activities are likely to differ between benign accounts and money- laundering accounts. In order to model the sequential behavior, we use the discrete-time Markov Chain model. Specifically, we record the sequence of three basic financial activities: virtual-currency recharge, self- expenditures, and expenditures as gifts. Each state in the Markov Chain corresponds to one activity and the transition between two states represents a pair of two consecutive finan- cial activities. Hence, the Markov Chain has three states and nine total transitions. Each transition is associated with the probability of this transition among all observed transitions. Figure 3a illustrates how Markov Chain models are derived from a sequence of financial activities. Specifically, nodes 1', 2', and 3' refer to the three states "virtual- currency exchange," "self-expenditure," and "expenditure as gifts"; P_{ij} denotes the transition probability from state i to state j.

Figure 3b presents the CDF of P_{11}, P_{31}, and P_{33} for malicious accounts (denoted as "MA") and benign accounts (denoted as "BA"), respectively. As shown in the empirical analysis, the values of P_{11} and P_{33} for malicious accounts are much larger than those for benign accounts, which indicates that malicious accounts are more inclined to exchange multiple times continuously (see P_{11}), and expend as gifts multiple times continuously (see P_{33}). The values of P_{31} of malicious accounts are much smaller than those of benign accounts, which implies that benign accounts are more active to recharge virtual currency after expending as gifts compared to malicious accounts. It is worth noting that we omit the other six transition probabilities in the figure for brevity.

Our empirical analysis demonstrates that the sequential behaviors indeed experience significant differences between malicious and benign accounts. Therefore, we define the following features.

FEATURES 9–17: The transition probabilities P_{11}, P12, P13, P21, P22, P23, P31, P32, P33.

FeatUres 18–47: The top 30 most effective subsequences mined from the sequence of financial activities for malicious accounts. To achieve an acceptable time complexity, the

ISBN 978-93-88808-61-3

PrefixSpan algorithm [11] is used to mine the frequent subse- quences of behavior sequences. Then, the effec- tiveness e of mined subsequence q is measured by Eq. 1. In the equation, f_q denotes the number of times that subsequence q occurs in all the is associated with the probability of this tran- sition among all observed transitions. Figure 3a illustrates how Markov Chain models are derived from a sequence of financial activities. Specifical- ly, nodes 1', 2', and 3' refer to the three states "virtual-currency exchange," "self-expenditure," and "expenditure as gifts"; P_{ij} denotes the transi- tion probability from state i to state j.

Figure 3b presents the CDF of P_{11}, P_{31}, and P_{33} for malicious accounts (denoted as "MA") and benign accounts (denoted as "BA"), respectively. As shown in the empirical analysis, the values of P_{11} and P_{33} for malicious accounts are much larger than those for benign accounts, which indicates that malicious accounts are more inclined to exchange multiple times continuously (see P_{11}), and expend as gifts multiple times continuously (see P_{33}). The values of P_{31} of malicious accounts are much smaller than those of benign accounts, which implies that benign accounts are more active to recharge virtual currency after expending as gifts compared to malicious accounts. It is worth noting that we omit the other six transition probabilities in the figure for brevity.

Our empirical analysis demonstrates that the sequential behaviors indeed experience significant differences between malicious and benign accounts. Therefore, we define the following features.

FEATURES 9–17: The transition probabilities P_{11}, P12, P13, P21, P22, P23, P31, P32, P33.

FeatUres 18–47: The top 30 most effective subsequences mined from the sequence of financial activities for malicious accounts. To achieve an acceptable time complexity, the PrefixSpan algorithm [11] is used to mine the frequent subsequences of behavior sequences. Then, the effectiveness e ofmined subsequence q ismeasured by Eq. 1. In the equation, f_q denotes the number of times that subsequence q occurs in all the spAtIAl FeAtures oF currency trAnsFer

Each transaction for currency transfer can be characterized as a tuple denoted as $<s, t>$, where s and t refer to the source and destination account, respectively. For a node s, we identify a set of nodes, to each of which s has transferred virtu- al currency. We denote this set of nodes as $D(s)$ for this node$s$.We then

one or a few accounts (e.g., as birthday gifts) and thus form a fully connected graph, whose edges, however, are likely to have small weights. Since an account may receive gifts from both benign accounts (e.g., friend accounts) and money-laundering accounts, edges that connect benign and money- laundering accounts will also exist. To summarize, the graph is mainly composed of three types of connected subgraphs: subgraphs entirely composed of fully- connected malicious accounts, subgraphs entirely composed of fully-connected benign accounts, and subgraphs composed of both malicious accounts and benign accounts. Figure 4a presents one example of the third type of connected subgraphs. Specifically, malicious accounts A-D and C-E transfer to the same destination accounts respectively, a destination account obtains the virtual currency from both malicious account E and benign account F, and benign accounts F-I transfer to the same desti- nation account. Then, the corresponding graph is a connected graph composed of both malicious accounts and benign accounts.

ISBN 978-93-88808-61-3

Through analyzing the behaviors of destina- tion accounts, we find that most of the destination accounts as buyers tend to purchase the virtual currency or goods from the launder accounts rather than receiving gifts from benign accounts, and other destination accounts behave in the oppo- site way. This finding is validated by analyzing the neighbors of malicious accounts and benign accounts in the graph. It is analyzed that 80.1 percent of the neighbors of malicious accounts are malicious and 84.3 percent of the neighbors of benign accounts are benign on average. Thus, the malicious accounts and benign accounts tend to connect with the same type of vertices, and form a community structure in which some dense- ly connected components are composed of the same type of vertices and the connections among the components are sparse. The interpretation of the forming of community structure in transferring related graph is shown in Fig. 4a, and a real illustration of the structure is shown in Fig. 4b, where the red vertex denotes a malicious account and the blue vertex denotes a benign account.

To design the spatial features, we process the behaviors of accounts in the following two steps.

Step 1: Form the graph based on the definition of $G(V, E)$.

Step 2: Detect the densely-connected sub- graphs (communities) of the connected subgraphs of G based on a widely-used community detec- tion method, Fast Unfolding [12]. The method is a heuristic method based on modularity optimization, and is capable of dealing with large weighted graphs due to its acceptable time complexity.

Following the above two steps, the graph G will be divided into many communities, each account will belong to a community, and each community will be composed of almost the same type (malicious or benign) of accounts. We pres- ent the features of each account (vertex) below.

FEATURES of General ATTRIBUTES of Vertex in Graph

FeatUre 48 — Degree: The number of con-

detectIon And evAluAtIon

We leverage machine learning techniques to inte- grate all these features toperform effective detec- tion. Specifically, feature values extracted from labeled malicious and benign users have been employed to train a statistical classifier. After an unknown user is represented by a vectorof fea-

ture values, the classifier can automatically eval-

In order to evaluate the effectiveness of the proposed detection method, we use a total number of 496,414 accounts, of which 114,891 are malicious and 381,523 are benign. Without the loss of generality, we use Support Vector Machine(SVM), Random Forest (RM), and Logistic Regression (LR) [14] as the statistical classifier, where the SVM classifier was trained with a Gaussian Kernel and the RF classifier was trained with 3000 trees. We use three metrics to quantify the effectiveness of our method: detection rate (same definition as the true positive rate), false positive rate (FPR), and the area under the ROC curve (AUC) [15]. Specifically, AUC is a widely-used measure of the quality of the statistical classifier. It is defined as the probability that a randomly chosen sample of malicious accounts will have a higher estimated probability of belonging to malicious accounts than that of benign accounts. Since AUC is cutoff-independent and the values of AUC range from 0.5 (no predictive ability) to 1.0 (perfect predictive ability), a higher AUC of a classifier indicates better prediction performance, irrespective of the cutoff selection.

We perform 10-fold cross-validation to evaluate the detection performance of each select- ed statistical classifier based on all features,

ISBN 978-93-88808-61-3

using metrics including DR, FPR, and AUC. The results are presented in Table 1. Both Support Vector Machine and Random Forest can achieve high detection rates, high AUC values, and very low false positive rates. These results demonstrate that the features we extract can effectively differentiate between malicious accounts and benign accounts.

We evaluate the effectiveness of our meth- od when using features from one aspect or two aspects. Table 1 presents the results when SVM is adopted as the statistical classifier. The experimen- tal results demonstrate that features from each aspect show great promise in effectively detecting malicious accounts; features of two aspects show better performance compared to features from one aspect; the integration of features from all three aspects show the best performance. This

TABLE 1. Performance analysis of the detection method.

Classifiers	Features	FPR	Detection rate	AUC
SVM	All features	0.97 %	94.2%	0.966
RF	All features	0.22	92.3	0.960
LR	All features	4.56	90.2	0.928
SVM	Vitality features	3.0	86.9	0.920
SVM	Sequential features	3.83 %	93.3%	0.947
SVM	Spatial features	2.4	91.6	0.946
SVM	Vitality + sequential features	1.47	92.9	0.957
SVM	Vitality + spatial features	1.64	93.7	0.961
SVM	Sequential + spatial features	1.38 %	94.0%	0.963

implies high robustness of the proposed method. Specifically, if features of one aspect are evaded by attackers, remaining features can still accom- plish high detection accuracy.

On the scalability of the proposed detection method, although some of the vitality features may not be suitable for all the social networks (e.g., Feature 3 — percentage of rechargefrom mobile payment, because of that not all the social networks support mobile payment), the sequential and spatial features can be extracted in almost all the social networks integrating vir- tual currency, and are effective enough to detect the malicious accounts according to the performance analysis shown in Table 1. Therefore, other social networks can also adopt and extend the proposed method to detect the money-laun- dering accounts.

We also analyze the contribution ofeach single feature using information gain, where a higher value of information gain indicates more signifi- cant contribution. The rank of each feature based on information gain is shown in Table 2, where the top 20 features consist of seven spatial fea- tures, eight sequential features, and five vitality features. This indicates that all three aspects are useful for detection.

Conclusions

This article presents the analysis and detection method of money-laundering accounts in OSNs. We analyzed and compared the behavior of both malicious and benign accounts from three perspectives: the account viability, the transaction sequences, and spatial correlation among accounts. We designed a collection of 54 features to systematically characterize the behavior of benign accounts

and malicious accounts. Experimental results based on labeled data collected from Tencent QQ, a global leading OSN, demonstrated that the proposed method achieved high detection rates and very lowfalse positive rates.

References

[1] Y. Wang and S. D. Mainwaring, "Human-Currency Interaction: Learning from Virtual Currency use in China," *Proc. SIGCHI Conf. HUMAN Factors in COMPUTING Systems*, ACM, 2008, pp. 25–28.

[2] Y. Zhou *et al*., "ProGuard: Detecting Malicious Accounts in Social-Network-Based Online Promotions," *IEEE Access*, vol. 5, 2017, pp. 1990–99.

[3] F. Wu *et al*., "Social Spammer and Spam Message Co-Detection in Micro blogging with Social Context Regularization," *Proc. 24th ACM Int'l. Conf. Information and Knowledge Man-agement*, ACM, 2015, pp. 1601–10.

[4] L. Wu *et al*., "Adaptive Spammer Detection with Sparse Group Modeling," *Proc. 11th Int'l. AAAI Conf. Web and Social Media*, AAAI, 2017, pp. 319–26.

[5] S. Fakhraei *et al*., "Collective Spammer Detection in Evolving Multi-Relational Social Networks," *Proc. 21st ACM SIGKDD Int'l. Conf. Knowledge Discovery and Data Mining*, ACM, 2015, pp. 1769–78.

[6] F. Hao *et al*., "Robust Spammer Detection in Microblogs: Leveraging User Carefulness," *ACM Trans. Intelligent Sys- tems and Technology*, vol. 8, no. 6, 2017, pp. 83:1–31.

[7] G. K. Palshikar, "Detecting Frauds and Money Laundering: A Tutorial," *Proc. Int'l. Conf. Big Data Analytics*, Springer, 2014, pp. 145–60.R. Dreewski, J. Sepielak and W. Filipkowski, "The Application of Social Network Analysis Algorithms in a System Support- ing Money Laundering Detection," *Information Sciences*, vol. 295, 2015, pp. 18–32.

[8] E. L. Paula *et al*., "Deep Learning Anomaly Detection as Sup- port Fraud Investigation in Brazilian Exports and Anti-Money Laundering," *2016 15th IEEE Int'l. Conf. Machine Learning and Applications (ICMLA*), Anaheim, CA, 2016, pp. 954–60.

[9] A. F. Colladon and E. Remondi, "Using Social Network Analysis to Prevent Money Laundering," *Expert Systems with Applications*, vol. 67, 2017, pp. 49–58.

[10] J. Pei *et al*., "Mining Sequential Patterns by Pattern-Growth: The PrefixSpan Approach," *IEEE Trans. Knowledge and Data Engineering*, vol. 16, no. 11, 2004, pp. 1424–40.

[11] M. E. J. Newman, "Communities, Modules and Large-Scale Structure in Networks," *NATURE Physics*, vol. 8, no. 1, 2012, pp. 25–31.

[12] R. Li *et al.*, "Finding Influential Communities in Massive Net- works," *The VLDB JOURNAL*, 2017.

[13] S. Rogers, and M. Girolami, *A First CoUrse in Machine Learning*, CRC Press, 2016.

[14] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and TECHNIQUES*, Elsevier, 2011.

ABOUT MRGI

Malla Reddy group of Institutions is one of the biggest conglomerates of hi-tech professional educational institutions in the state of Telangana, established in 2001 sprawling over 200 acres of land. The group is dedicated to impart quality professional education like pharmacy, Engineering & Technology, MCA, MBA courses. Our sole objective is to turn out high caliber professionals from those students who join us.

ABOUT MRCE

Malla Reddy group of Engineering (Formerly CM Engineering College) has been established under the aegis of the Malla Reddy Group of Institutions in the year 2005, a majestic empire, founded by chairman Sri Ch.Malla Reddy Garu. He has been in the field of education for the last 23 years with the intention of spearheading quality education among children from the school level itself. Malla Reddy College of Engineering has been laid upon a very strong foundation and has ever since been excelling in every aspect. The bricks of this able institute are certainly the adept management, the experienced faculty, the selfless non-teaching staff and of course the students.

ABOUT ICTIMES

ICTIMES started long back with its banner to promote the vision of future technologies that change the trends of life on this planet earth. Under this banner, the Department of Computer Science and Engineering at MRCE organizes the ICETCS – International Conference on Emerging Technologies in Computer Science to provide a scholarly platform to ignite the spirit of Research and bring out the latent potential in teaching fraternity and student community. ICETCS accommodates major areas like, Big Data , Data mining, Information Retrival, Neural Networks, Data Security, etc.,

ABOUT ICETCS

International Conference on Emerging Technologies in Computer Science ((ICETCS -2019) will bring together innovative academicians, researchers and industrial experts in the field of Computer Science to a common forum. The idea of the conference is for the scientists, scholars, engineers and students from the Universities across the world and the industry as well, to present ongoing research activities, and hence to foster research relations between the Universities and the industry with the rapid development of trends and studies in the fields concerned. ICETCS-2019 will provide a heartwarming platform to researchers, scholars, faculty and students to exchange their novel ideas face to face togethers.

Estd :2005

MALLA REDDY COLLEGE OF ENGINEERING

Approved by AICTE - New Delhi, Affiliated to JNTU - Hyderabad,Accredited by NBA & Accredited by NAAC.
ISO 9001:2015 Certified Institution, Recognition of College under Section 2(f) & 12 (B) of the UGC Act,1956.

Address: Maisammaguda, Dhulapally,(Post Via Kompally),Secunderabad - 500 100.

Ph: 040-64632248, 9348161222 ,9346162620. Email: principal@mrce.in

Website : www.mrce.in

ICTIMES 2019

5th INTERNATIONAL CONFERENCE ON

Trends in Information, Management, Engineering and Sciences

Proceedings of International Conference on Latest Trends in Electronics and Communication (ICLTEC)

ISBN No: 978-93-88808-62-0

Convenor
Dr. R. P. Naik
HOD-Dept. of ECE

Editor
Dr. B BALAJI
Professor - ECE

Editor in Chief
Dr. P. John Paul
Principal

Organized by

Estd :2005

MALLA REDDY COLLEGE OF ENGINEERING

Approved by AICTE - New Delhi, Affiliated to JNTUH, Accredited by NBA & Accredited by NAAC.
ISO 9001:2015 Certified Institution, Recognition of College under Section 2(f) & 12 (B) of the UGC Act,1956.

website: www.mrce.in

Proceedings on International Conference on Electronics and Communication (ICLTEC-2019)
ISBN 978-93-88808-62-0

ICTIMES - INTERNATIONAL CONFERENCE ON ELECTRONICS AND COMMUNICATION (ICLTEC-2019)

ISBN 978-93-88808-62-0

INVENTIONS , INNOVATIONS AND STARTUPS IN HIGHER EDUCATIONAL INSTITUTIONS

Dr.John Paul Pulipati,Principal ,principal@mrce.in,Malla Reddy College of Engineering,Hyderabad,India

Dr.Kezia Rani Burgula,Associate Professor,CSE Department,keziarani@stanley.edu.in,Stanley College of Engineering for Women,Hyderabad,India

ABSTRACT

The paper advances current knowledge on factors affecting higher education institutions in their quest for innovation in education. This paper takes an innovative approach by adopting the concept of 'innovation systems' and adapting it for higher education. The analytical construct of 'higher education innovation system' has thus been developed as a sub-set of an innovation system, concentrated particularly in higher education institutions which are close connection with other institutional spheres, such as industry, government and non-government agencies, and the society at large in terms of Internship and innovation process. As the part of the interns discussed with classification internship and also how it has been played major role in the institution towards to the Higher Education process. In higher education innovation system has to be following with set of functions, components and relationships, which allow us to disaggregate the various levels of interactions among the elements of the system and analyse the unfolding of innovation in higher education. Results indicate certain "disengagement" in relation of higher education institutions and education policy makers, business, and students as well as between higher education institutions' managers and their subordinates. Based on the findings, major innovation, incubation and startup related challenges in the higher education are discussed and related practical recommendations are presented.

Key words: Innovation, start up, intellectual property rights, entrepreneurship, technology transfer, technology commercialization, technology policy, university-industry links, economic development, R&D, knowledge management, incubation

1. INTRODUCTION

Higher education is changing across Indian and other global countries and there is a growing expectation from policy makers and society that higher education institutions (HEIs) should

evolve into a new type of economic actor. Entrepreneurship and innovation in higher education are no longer only associated with business start-ups and technology transfer but are increasingly understood as core elements of a procedural framework for how organisations and individuals behave. For example, in how links between teaching and research are created and nurtured, how societal engagement and knowledge exchange are organized, how resources are built and managed for effective partnerships, and how new entrepreneurs are supported [1].

One of the major contributions in fields of application and impact of innovation, incubation and internship in HEIs has been laid forward by William Spady, a sociologist and the father of Outcome-Based Education (OBE). OBE is referred to by over 20 different names including Systemic education restructuring, Performance Based Education, Standards based education reform, High Performance Learning, Total Quality Management, Transformational Education, and Competency-Based Education [8-14].Outcome-Based Education[2-6] means clearly focusing and organizing everything in an educational system around what is essential for all students to be able to do successfully at the end of their learning experiences. This means starting with a clear picture of what is important for students to be able to do, then organizing curriculum, instruction, and assessment to make sure this learning ultimately happens. The keys to having an outcome-based system are [9-10]:

1. Developing a cleat set of learning outcomes around which all of the system's components can be focused.
2. Establishing the conditions and opportunities within the system that enable and encourage all students to achieve those essential outcomes.

For example, the possible outcome "explain the major causes of inflation in capitalist economies" implies that to be successful, the learner will be expected to develop both the competence of explaining and knowledge of major causes of inflation in capitalist economies. Since outcome-based systems expect learners to earn out the processes defined within an outcome statement, they are careful to build those processes directly into the outcome through demonstration verbs. Therefore, one key to recognizing a well-defined outcome is to look for the demonstration verb or verbs that define which processes the learner is expected to carry out at the end. Without those verbs, what are called outcome statements lack a clearly defined demonstration process, and without that defined process the outcome statement takes on the character of a goal rather than a true outcome demonstration [11-12].

1.2 Program Educational Objectives (PEO)

Graduates are able to:

1.2.1 PEO-1: Work in automotive and related industries by applying the knowledge of science, mathematics and engineering.

1.2.2 PEO-2 Demonstrate team work and project management skills in a multi-disciplinary environment through effective communication, modern tool usage and professional ethics.

1.2.3 PEO-3 Investigate, analyse, research and solve problems in the field of automotive planning, designing, manufacturing, testing and servicing.
1.2.4 PEO-4 Pursue lifelong learning through professional trainings and practices with consideration of safety, environment and sustainability

Following are the Program Learning Outcomes (PLOs) adopted in Automotive Engineering Program as recommended by PEC in their Manual of Accreditation 2014.

1.2.5 PLO-1 Engineering Knowledge: An ability to apply knowledge of mathematics, science, engineering fundamentals and an engineering specialization to the solution of complex engineering problems.
1.2.6 PLO-2 Problem Analysis: An ability to identify, formulate, research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences and engineering sciences.
1.2.7 PLO-3 Design/Development of Solutions: An ability to design solutions for complex engineering problems and design systems, components or processes that meet specified needs with appropriate consideration for public health and safety, cultural, societal, and environmental considerations.[3,4]
1.2.8 PLO-4 Investigation: An ability to investigate complex engineering problems in a methodical way including literature survey, design and conduct of experiments, analysis and interpretation of experimental data, and synthesis of information to derive valid conclusions.
1.2.9 PLO-5 Modern Tool Usage: An ability to create, select and apply appropriate techniques, resources, and modern engineering and IT tools, including prediction and modeling, to complex engineering activities, with an understanding of the limitations.

1.32.10 PLO-6 The Engineer and Society: An ability to apply reasoning informed by contextual knowledge to assess societal, health, safety, legal and cultural issues and the responsibilities relevant to professional engineering practice and solution to complex engineering problems.
1.2.11 PLO-7 Environment and Sustainability: An ability to understand the impact of professional engineering solutions in societal and environmental contexts and demonstrate knowledge of and need for sustainable development.
1.2.12 PLO-8 Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of engineering practice.
1.2.13 PLO-9 Individual and Team Work: An ability to work effectively, as an individual or in a team, on multifaceted and /or multidisciplinary settings.
1.2.14 PLO-10 Communication: An ability to communicate effectively, orally as well as in writing, on complex engineering activities with the engineering community and with society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
1.2.15 PLO-11 Project Management: An ability to demonstrate management skills and apply engineering principles to one's own work, as a member and/or leader in a team, to manage projects in a multidisciplinary environment.

1.2.16 PLO-12 Lifelong Learning: An ability to recognize importance of, and pursue lifelong learning in the broader context of innovation and technological developments.

1.3 A S K (Attitude Skill Knowledge)

Attitude, skill and knowledge[1,2], all the three elements put together form a success formula. It is a well-known fact that any person with a remarkable blend of all the above mentioned 3 characteristics will come out as a human being with lots of intelligence, leadership qualities and mentoring abilities. But, the question is, Can any one of these attributes stand alone and guide a person to the zenith? The answer is probably NO. As, we all know and understand that in the present era of competitiveness and naivety every individual is expected to be skillful, knowledgeable and also must possess a correct set of attitude which is universally acceptable. A person with only a bucket full of skills required to complete a task with a very less amount of knowledge on the subject and probably with a poor attitude cannot be considered useful for an organization for the long run. If you don't have skills, it can be acquired. If you don't have knowledge, it can be gained. But, if you don't have attitude, you are in trouble. Attitude, in a general sense is taken negatively. The triangle of success represented by ASK is shown in Fig. 1 [21-23].

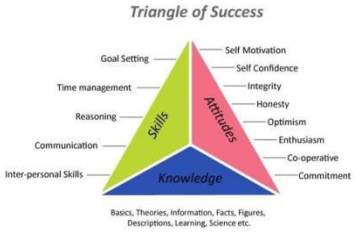


Fig. 1 Attitude, skill and knowledge- The triangle of success

1.4 Blooms Taxonomy

Bloom's taxonomy is a set of three hierarchical models used to classify educational learning objectives into levels of complexity and specificity. The three lists cover the learning objectives in cognitive, affective and sensory domains. Various stages of Blooms taxonomy is shown in Fig. 3 [13-18].

1.3.2 The Revised Taxonomy (2001)

A group of cognitive psychologists, curriculum theorists and instructional researchers, and testing and assessment specialists published in 2001 a revision of Bloom's Taxonomy with the title *A Taxonomy for Teaching, Learning, and Assessment*. This title draws attention away from the somewhat static notion of "educational objectives" (in Bloom's original title) and points to a more dynamic conception of classification.

- Remember
  - Recognizing
  - Recalling
- Understand
  - Interpreting
  - Exemplifying
  - Classifying
  - Summarizing

- Inferring
- Comparing
- Explaining

- Apply
  - Executing
  - Implementing
- Analyze
  - Differentiating
  - Organizing
  - Attributing
- Evaluate
  - Checking
  - Critiquing
- Create
  - Generating
  - Planning
  - Producing

Bloom's revised taxonomy is shown in Fig. 2.

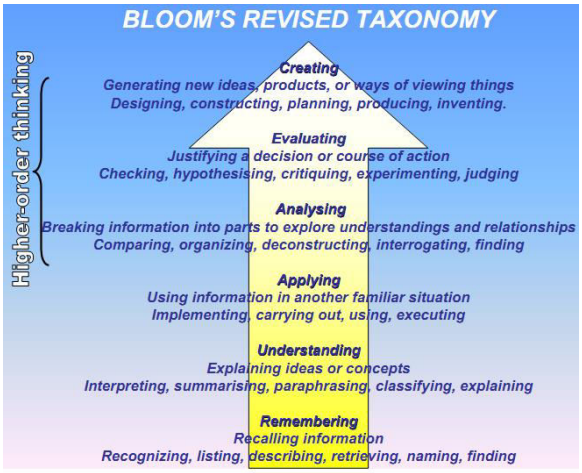


Fig. 2 Bloom's revised taxonomy

II Internship:

An internship is an official program offered by an employer to potential employees. Interns work either part time or full time at a company for a certain period of time. Internships are most popular with undergraduates or graduate students who work between one to four months and have a goal to gain practical work or research related experience. The main difference between an apprenticeship and an internship is that internships are more exploratory. You're not bound to work for your employer after the internship is over (although many interns do receive job offers). If you start early enough to do a few internships throughout college, you can use the first ones to get a feel for what career you'd like to pursue and the later ones to build your experience. Internships can be paid or unpaid. For example, in California, unpaid interns must receive college credit for their work.

2.1 Benefits for Internship Program

1. Tech-savvy. The first generation to be brought up with computers, members of Gen Y are "digital natives." In fact, not only can they uncover, operate, and recommend the most advanced

technologies; they can teach you how to use tools like content management systems and social media.

2. Cost-effective. Compared to other populations, Generation Y appears less motivated by money. A study by UNC's Kenan-Flagler Business School[4] found that millennial prioritize meaningful work over higher pay.

3. Intrinsically motivated. As the results of the Kenan-Flagler Business School study demonstrate, millennial want to grow, develop, and advance within the workplace. They are intrinsically motivated to succeed, a quality all employers seek in their employees.

4. Team players. If some called Generation X "The Me Generation," we might term Generation Y "The We Generation" for their heightened sense of community and peer-to-peer relationships.

5. Highly educated. Millennials are on track to be the most educated generation yet. According to a 2016 study by the Pew Research Center, 20 percent of millennials were college graduates, with an additional 40 percent still in school. About half of millennials still in school planned to earn a graduate or professional school degree, and many millennials who had graduated planned to return.

6. Optimistic. Positivity in the workplace is invaluable to company culture and team morale, just as negativity can be extremely detrimental.

7. Current. Like employing any youth generation, hiring millennials helps keep your company up to date with social, entertainment, and other market trends. Millennials offer a fresh perspective and can generate marketing strategies that appeal to younger generations of consumers.

2.3 Role of Internship coordinator—this person generally works in the host organization's HR department and can be seen as the overall director of the internship program. The internship coordinator is usually responsible for overseeing the development and implementation of an internship program, overseeing and coordinating the program once it's up and running, assigning interns to various departments, and managing site supervisors. This individual also acts as the liaison between the company and the educational institution (the student's faculty sponsor, the school's career centre director, and/or career counselors).

In general, there are three aspects necessary for an internship to constitute a learning experience:

1. The internship is within the student's area of study; and they bring to the internship the knowledge they've acquired through their academic education.
2. The internship supervisor provides guidance, evaluation, and feedback to facilitate the learning process. (Evaluations and feedback, however, can be written or verbal.)
3. The student engages in ongoing contemplation of learning objectives throughout the course of the internship.

2.4 INTERNATIONAL INTERNSHIP: An International Development internship with Projects Abroad and gain skills that you will use for the rest of your career, regardless of what field you move into. As you get hands-on experience in the development field you'll learn about infrastructure and disaster management as well as aid and development in disadvantaged communities. You'll also gain a deeper understanding of the world and the complex processes that are involved in supporting people and places to grow and prosper. We currently have a number of International Development Projects, working to transform communities that need help. Your work could include creating campaigns, educating local communities, and facilitating existing projects. You'll work with local experts and Projects Abroad staff to better understand global development challenges and the work needed to overcome them. There are two main aims to this project: providing you with a practical, insightful work experience, while helping to address issues affecting development in Abroad (Developed Country).Many of the interns who join us are looking for hands-on experience in their field of study. We provide you with the opportunity to learn from skilled local professionals and gain first-hand insight into the issues facing. By sharing knowledge and experiences, you'll participate in valuable cross-cultural exchange. This will help prepare you for your future career. ***The goal of international development is to improve the wellbeing of local communities by working on sustainable solutions to their problems***. On this internship, you can choose to focus on a range of different issues, from immigration to human rights. Working with local NGOs who have expert knowledge of the situation, you'll contribute towards their efforts in protecting and empowering vulnerable groups of people. Help to create long-term positive change in local communities on this internship, while gaining invaluable international development work experience in Mexico.

3. Business incubator: is a company that helps new and startup companies to develop by providing services such as management training or office space. The National Business Incubation Association (NBIA) defines business incubators as a catalyst tool for either regional or national economic development. NBIA categorizes their members' incubators by the following five incubator types: academic institutions; non-profit development corporations; for-profit property development ventures; venture capital firms, and combination of the above [1-2].Most research and technology parks do not offer business assistance services, which are the hallmark of a business incubation program. However, many research and technology parks house incubation programs [3].Incubators also differ from the U.S. Small Business Administration's Small Business Development Centers (and similar business support programs) in that they serve only selected clients. SBDCs are required by law to offer general business assistance to any company that contacts them for help [4-5]. Within European Union countries

there are different EU and state funded programs that offer support in form of consulting, mentoring, prototype creation and other services and co-funding for them. TecHub is one of examples for IT companies and ideas [6].In India, the business incubators are promoted in a varied fashion: as Technology Business Incubators (TBI) and as Startup Incubators -- the first deals with technology business (mostly, consultancy and promoting technology related businesses) and the later deals with promoting startups (with more emphasis on establishing new companies, scaling the businesses, prototyping, patenting, and so forth).

Types of services

Since startup companies lack many resources, experience and networks, incubators provide services which helps them get through initial hurdles in starting up a business. These hurdles include space, funding, legal, accounting, computer services and other prerequisites to running the business [17-20].

Among the most common incubator services are:

Help with business basics

- Networking activities
- Marketing assistance
- Market Research
- Access to angel investors or venture capital
- Comprehensive business training programs
- Advisory boards and mentors
- Management team identification
- Help with business etiquette
- Technology commercialization assistance
- Intellectual property management

Types of Business Incubation :
There are a number of business incubators that have focused on particular industries or on a particular business model, earning them their own name.
This list is incomplete; you can help by expanding it.

- Virtual business incubator - online business incubator
- Kitchen incubator - a business incubator focused on the food industry
- Public incubator - a business incubator focused on the public good
- Seed accelerator - a business incubator focused on early startups

Corporate accelerator - a program of a larger company that acts akin to a seed accelerator

- Startup studio - a business incubator with interacting portfolio companies
- Technology Business Incubator - a business incubator based out academic institutions

- Hybrid Incubator - A business incubator that combines virtual incubator with on-premise activities

A *startup* is a young company founded by one or more entrepreneurs in order to develop a unique product or service and bring it to market. By its nature, the typical *startup* ends to be a shoestring operation, with initial funding from the founders or their families.

A startup venture could be defined as a new business that is in the initial stages of operation, beginning to grow and is typically financed by an individual or small group of individuals. It is a young entrepreneurial, scalable business model built on technology and innovation wherein the founders develop a product or service for which they foresee demand through disruption of existing or by creating entirely new markets. Startups[25-28] are nothing but an idea that manifests into a commercial undertaking.

Grant Thornton (2016) define startup business[26,28] as an organization which is

a) Incorporated for three years or less
a) At a funding stage of Series B or less(B Series means second round of funding)
b) An entrepreneurial venture/a partnership or a temporary business organisation
c) Engages in development, production or distribution of new products/services or processes
d) Revenue of up to INR 25 cr.
e) Not formed through splitting or restructuring
f) Employing 50 people or less

Department of Industrial Policy and Promotion (DIPP) define a startup as an entity incorporated or registered in India with following parameters:

- Established not prior to seven years, (for Biotechnology Startups not prior to ten years)
- With annual turnover not exceeding INR 25Cr in any preceding financial year, and
- Working towards innovation, development or improvement of products or processes or services,
- It is a scalable business model with a high potential of employment generation or wealth creation

It is to be noted that such entity is not formed by splitting up, or reconstruction, of a business already in existence. Also, an entity shall cease to be a startup if its turnover for the previous financial years has exceeded INR 25cr or it has completed 7 years (biotechnology startups 10 years) from the date of incorporation/ registration('Startup India', 2017)

GOVERNMENT INITIATIVES

Indian government is serious in promoting entrepreneurship at the startup level and has taken a number of initiatives to ensure appropriate support. In this aspect it is relevant to mention **'Make in India'** campaign introduced in September'14 to attract foreign investments and encourage domestic companies to participate in the manufacturing sector. The government increased the foreign direct investment (FDI) limits for most of the sectors and strengthened intellectual property rights (IPRs) protection to instill confidence in the startups.In order to make the country as number one destination for startups, Government of India (GoI) has introduced a new campaign called '**Standup India'** in 2015 aimed at promoting entrepreneurship among women and to help startups with bank funding. Another commendable and far reaching initiative is '**Digital India'** introduced in 2015 to ensure government services are made available to every citizen through online platform that aims to connect rural areas by developing their digital infrastructure which translates into a huge business opportunity for startups.

THE STARTUP SCENARIO IN INDIA

It is to be noted that every year more than 800 technology startups are being set up in India. By 2020, it is estimated that around 11,500 tech-startups are going to be established with employment potential of around 250,000 technical people (NASSCOM, 2015). It is admirable to note that India is amongst the top five countries in the world in terms of startups with 10,000+ led by US with 83,000+ comprising 43% tech-based firms with 9% managed by women entrepreneurs. The number of incubators also has crossed 100 in 2014-15 to give boost to the startup saga (Grant Thornton, 2015).Sector wise, the distribution of Indian businesses is:

Table:1 Break-up of Indian Startup Businesses

Technology Based	Non-Technology Based
E-Commerce - 33%	Engineering- 17%
B2B - 24%	Construction-13%
Internet - 12%	Agri- products- 11%
Mobile apps - 10%	Textile - 8%
SaaS - 8%	Printing & packaging – 8%
Other – 13%	Transport & logistics- 6%
	Outsourcing & support -5%
	Others-32%

Source: Startups India- An Overview, Grant Thornton, 2015

THE STARTUP ECOSYSTEM

Along with government initiatives, there is a definite movement in startup arena in India due to penetration of IT and internet. Many startups[30-34] are coming up in service sector including education, legal, retail, insurance and health. With customers becoming aware of the benefits and convenience, the popularity and viability of startups is no more a difficult proposition for an entrepreneur.

A number of venture capitalists and angel investors are aggressive and gung-ho on Indian startups as they see lot of potential with few expected to become unicorns (high valued companies) bringing in good returns. On the contrary, there are examples of few startups that failed and eventually closed their businesses due to various issues and challenges.

ISSUES AND CHALLENGES OF STARTUPS

A successful start-up cannot start a business just with passion and an idea. A high level of leadership skills with clear understanding of market, excellent communication skills, maturity to see things in right perspective along with the ability to take calculated risks are required on the part of the entrepreneur(Aggarwal,2017). Lack of awareness, multiple clearances, unorganised market, poor infrastructure in Tier 2 /3 cities, lack of mentoring , stringent exit policies, corruption/red tape, technological risk, regulatory obstacles and lack of reforms keeping pace with the fast evolving market changes are some of the challenges as per Rashmi Guptey, Principal (Legal) of Lightbox India Advisors Private Limited.

Some of the major issues and challenges are discussed below:

Financial Resources

Availability of finance is critical for the startups and is always a problem to get sufficient amounts (Mittal, 2014; Truong, 2016). A number of finance options ranging from family members, friends, loans, grants, angel funding, venture capitalists, crowdfunding etc are available. The requirement starts increasing as the business progresses. Scaling of business requires timely infusion of capital. Proper cash management is critical for the success of the

startups (Skok, 2016;Pandita,2017). A recent report paints a gloomy picture with 85% of new company's reportedly underfunded indicating potential failure (Iwasiuk, 2016).

Revenue Generation

Several startups fail due to poor revenue generation as the business grows. As the operations increase, expenses grow with reduced revenues forcing startups to concentrate on the funding aspect, thus, diluting the focus on the fundamentals of business. Hence, revenue generation is critical, warranting efficient management of burn rate which in common parlance is the rate at which startups spend money in the initial stages. The challenge is not to generate enough capital but also to expand and sustain the growth.

1) Team Members

To find and hire the right kind of talent for the business with skills to match growing customer expectations are one of the biggest challenges (Truong,2016). Apart from founder(s), startups normally start with a team consisting of trusted members with complementary skill sets. Usually, each member is specialized in a specific area of operations. Assembling a good team is the first major requirement, failure to have one sometimes could break the startup (Skok, 2016). According to a survey, 23 percent startups failed because members did not work as a team. Chirag Garg, CEO, HyperDell, feels that **bringing in affordable talent at the right time is a challenge. As per** Nitin Sharma, Principal & Founding member, Lightbox India Advisors Private Limited "Hiring and retaining high quality talent, especially in the areas of product and technology remains a key challenge" (Choudhary,2015)

2) Supporting Infrastructure

There are a number of support mechanisms that play a significant role in the lifecycle of startups which include incubators, science and technology parks, business development centers etc. Lack of access to such support mechanisms increases the risk of failure.

3) Creating Awareness in Markets

Startups fail due to lack of attention to limitations in the markets. The environment for a startup is usually more difficult than for an established firm due to uniqueness of the product. The situation is more difficult for a new product as the startup has to build everything from scratch.

4) Exceed Customer Expectations

The next most important challenge is gauging the market need for the product, existing trends, etc. Innovation plays an important role, since, that the startup has to fine-tune the product offerings to suit the market demands (Skok, 2016). Also, the entrepreneur should have thorough domain knowledge to counter competition with appropriate strategies. Due to new technologies that are emerging, the challenge to provide over and above an earlier innovation is pertinent. Namrata Garg, Director, SendKardo feels that t**he biggest challenge is the need to constantly reinvent yourself and come up with a service to be able to match up customer expectations** and exceed them.

5) Tenacity of Founders

Founders of startups have to be tough when the going gets tough. The journey of starting a venture is fraught with delays, setbacks and problems without adequate solutions. The entrepreneur needs to be persistent, persuasive, and should never give up till he/she achieves desired results. History is replete with startups who gave up the fight when things went wrong. Sometimes the product could be ahead of its time or may require complimentary technology /products for the use by the customers. For example, Apple had to delay introduction of iTunes till the regulations favoured the launch. It is also relevant to quote Steve Jobs who by

commenting "*A lot of times, people don't know what they want until you show it to them"* reiterates the fact those products from startups mostly fall in the "new and untried" category where the success rate is minimal.

6) Regulations

Starting a business requires a number of permissions from government agencies. Although there is a perceptible change, it is still a challenge to register a company. Regulations pertaining to labor laws, intellectual property rights, dispute resolution etc. are rigorous in India which takes about 30 days to comply compared to just 9 days in OECD countries. Also, as per World Bank report, "World Bank Ease of Doing Business", India ranks 142 out of 189 economies (Mittal,2014).

7) Growth Decelerators

Some of the agencies which are part of the startup ecosystem themselves can sometimes become hurdles in the growing stages. As per Sneh Bhavsar, co-founder and CEO, OoWomaniya one of the major issues is **the influence of** incubators, institutes and similar organisations which try to control, manage and be the daddies of the start-ups in the name of helping, mentoring etc (Choudhury,2015).This needs proper coordination among the organizations for mutual benefit.

8) Lack of Mentorship

Milan Hoogan, Vice President -Sales and Marketing at Erfolg Life Sciences feels that l**ack of proper guidance and mentorship** is one of the biggest problems that exist in the Indian startup ecosystem (Choudhury, 2015) . Most of startups have brilliant ideas and/or products, but have little or no industry, business and market experience to get the products to the market. It is a proven example that a brilliant idea works only if executed promptly (Mittal,2014). Lack of adequate mentoring/guidance is the biggest challenge which could bring a potentially good idea to an end.

9) Lack of a Good Branding Strategy

Absence of an effective branding strategy is another issue that prevents startups from flourishing at a faster pace. Hemant Arora, Business Head-Branded Content, Times Network opines that branding demands paramount attention as it gives an identity and occupies a space in the consumer minds(Choudhury,2015) .

10) Replicating Silicon Valley

Koushik Shee, Founder and CEO, Effia **,** feels that Indian startups get influenced by Silicon Valley models which may not succeed in Indian scenario. Lot of tweaking and modifications could be required when transplanted into Indian markets keeping in mind Indian infrastructure in terms of roads, internet, electricity and telecom penetration (Choudhury,2015) .

1. REASONS FOR FAILURE

As regards major reasons for failure of startups, a survey based on analysis of 101 firms showed that 42% failed as the product had no market, 29% firms ran out of cash, 23% did not have the right team,18% closed due to pricing issues, 17% firms had poor product, 14% failed due to poor marketing and 8% had no investor interest(Griffith,2014). These reasons substantiate most of the issues and challenges that have been enumerated above.

GOVERNMENT INITIATIVES

There are numerous government and semi-governmental initiatives to assist startups.

- **Start-Up India**

  This initiative provides three-year tax and compliance breaks intended for cutting government regulations and red tapism.

- *MUDRA Yojna*

Through this scheme, startups get loans from the banks to set up, grow and stabilize their businesses.

- **SETU (Self-Employment and Talent Utilization) Fund**

Government has allotted Rs 1,000 Cr in order to create opportunities for self-employment and new jobs mainly in technology-driven domains.

- **E-Biz Portal**

Government launched e-biz portal that integrates 14 regulatory permissions and licenses at one source to enable faster clearances and improve the ease of doing business in India.

- **Royalty Tax**

Indian government has reduced the royalty tax paid by businesses and startup firms from 25per cent to 10 per cent.

CONCLUSIONS

In this paper we discussed about intership, innovation, incubation and startup with their types with national, international. Our proposed method describes the above parameters for technical institutions in india.The current economic scenario in India is on expansion mode. The Indian government is increasingly showing greater enthusiasm to increase the rate of growth from grass root levels with introduction of liberal policies and initiatives for entrepreneurs like 'Make in India', 'Startup India', MUDRA etc. 'Make in India' is great opportunity for the Indian start-ups. With government going full hog on developing entrepreneurs, it could arrest brain drain and provide an environment to improve availability of local talent for hiring by startup firms. In this paper,as the part of higher education discussion about types of internship to educate the engineers to the suitable working environment in the institutions. Those students has participated interested to learn and digest working environment with academics. while in the internship period the has learn and getting the idea generation to produce the product out come. Small contributions from a number of entrepreneurs would have cascading effect on the economy and employment generation which would complement medium and large industries efforts catapulting India into a fast growing economy. The startup arena has lot of challenges ranging from finance to human resources and from launch to sustaining the growth with tenacity. Being a country with large population, the plethora of opportunities available are many for startups offering products and services ranging from food, retail, and hygiene to solar and IT applications for day to day problems which could be delivered at affordable prices. It is not out of place to mention that some of these startups would become unicorns and may become world renowned businesses by expanding into other developing and underdeveloped countries.

REFERENCES

1. Brown, S., Bucklow, C., Clark, P., 2002. Professionalizing teaching: enhancing the status of teaching, improving the experience of learning and supporting innovation in higher education. J. Geogr. High. Educ. 26 (2), 159–168.

2. Amason, A. 1996. Distinguishing the effect of functional and dysfunctional conflict on strategic decision making: Resolving a Paradox for top management teams. Academy of Management Journal, 39 pp. 123-148.
3. Bichsel, J. 2012. Analytics in Higher Education: Benefits, Barriers, Progress, and Recommendations. [e-book] EDUCAUSE Center for Applied Research. Available at: http://www.educause.edu/eca.
4. Bohle Carbonell, K. et al. 2012. Unleashing the creative potential of faculty to create blended learning. The Internet and Higher Education.
5. Brennan, J. et al. 2013. Things we know and don't know about the Wider Benefits of Higher Education: A Review of the Recent Literature. BIS Research Paper Number 133.
6. Venkatesh, U., & Dutta, K. (2007). Balanced scorecard in managing higher education institutions: An India perspective. International Journal of Educational Management, 21(1), 54–67.
7. Vishnoi, A. (2013). Innovation Universities Bill Stuck, University Grants Commission Looks at Giving Funds to Existing Universities for Innovation, The Indian Express. Retrieved from http://www.indianexpress.com/news/innovation-universities-bill-stuck-universi/1068853/.
8. Spady, William (1994). *Outcome-Based Education: Critical Issues and Answers*(PDF). Arlington Virginia: American Association of School Administrators. ISBN 0876521839. Retrieved 31 October 2014.
9. Mohayidin, Mohd Ghazali (2008). "Implementation of Outcome-Based Education in Universiti Putra Malaysia: A Focus on Students' Learning Outcomes". *International Education Studies*. 1 (4). Retrieved 23 October 2014.
10. Kaliannan, Maniam; Chandran, Suseela Devi (2012). "Empowering Students through Outcome-Based Education (OBE)". Research in Education. 87 (1): 50–63.
11. Desai, S., & Kurkarni, V. (2008). Changing educational Inequalities in India in the context of affirmative action. Demography, 45(2), 245–270.
12. Dube, S. (1988). Higher education and social change. In A. Singh & G. Sharma (Eds.), Higher education in India: The social context (pp. 120–176). New Delhi: Konark.
13. Bloom, B. S. (1994). "Reflections on the development and use of the taxonomy". In Rehage, Kenneth J.; Anderson, Lorin W., Sosniak, Lauren A. (eds.). *Bloom's taxonomy: A forty-year retrospective*. Yearbook of the National Society for the Study of Education. **93**. Chicago: National Society for the Study of Education. ISSN 1744-7984.
14. http://www.kurwongbss.qld.edu.au/thinking/Bloom/blooms.htm.
15. Anderson, Lorin W. & Krathwohl, David R. (2001). A Taxonomy for Learning, Teaching and Assessing: a Revision of Bloom's Taxonomy. New York. Longman Publishing.
16. Clark, Donald R. (1999). "Bloom's Taxonomy of Learning Domains". Retrieved 28 Jan 2014.

17. Hoy, Anita Woolfolk (2007). Educational psychology (10th ed.). Boston: Pearson/Allyn and Bacon. pp. 530–531, 545. ISBN 0205459463. OCLC 68694368.
18. Armstrong, Patricia. "Bloom's Taxonomy". Vanderbilt University Center for Teaching. Vanderbilt University. Retrieved 29 June 2016.
19. Harrow, Anita J. (1972). A taxonomy of the psychomotor domain: A guide for developing behavioral objectives. New York: David McKay Company.
20. Dave, R. H. (1975). Armstrong, R. J. (ed.). Developing and writing behavioral objectives. Tucson: Educational Innovators Press.
21. M. Yang, M. You and F. Chen "Competencies and qualifications for industrial design jobs: implications for design practice, education, and student career guidance" Design Studies, Vol. 26, Elsevier Ltd, UK, (2005), 155-189.
22. W. Lewis, and E. Bonollo, "An analysis of professional skills in design: implications for education and research", Design Studies, Vol. 23, Elsevier Ltd, UK, (2002), 385-406.
23. A. Salama, "A Voice for an alternative architectural education: integrating "What" and "How" knowledge" ARCHITTIMES, Karachi, Pakistan, September (2003).
24. Agarwal, P. (2009). Indian higher education: Envision the future. New Delhi: SAGE.
25. Brown, P., Lauder, H., & Ashton, D. (2010). The global auction—The broken promises of education, jobs, and incomes. Oxford: Oxford University Press.
26. Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). The new production of knowledge: The dynamics of science and research in contemporary societies. London: SAGE.
27. Government of India. (1958). Scientific policy resolution. In Government of India. Department of
28. Science and Technology (Ed.). New Delhi: Ministry of Science and Technology.
29. Government of India. (1983). Technology policy statement. In Department of Science & Technology (Ed.). New Delhi: Ministry of Science and Technology.
30. Government of India. (2003). Science and technology policy. In Department of Science & Technology (Ed.). New Delhi: Ministry of Science and Technology.
31. Government of India. (2012). Twelfth five year plan (2012–2017) social sectors. In Planning Commission (Ed.), (Vol. 3, p. 274). New Delhi: SAGE.
32. Government of India. (2013). Science, technology and Innovation policy. In Minister of Science & Technology (Ed.), (p. 22). New Delhi: Ministry of Science and Technology.
33. Bound, K., & Thornton, I. W. (2012). Our frugal future: Lessons from India's innovation system. London: NESTA.
34. Kaul, S. (2006). Higher Education in India: Seizing the Opportunity Working Paper Series (Vol. 179). New Delhi: Indian Council for Research on International Economic Relations.

Low Overhead Warning Flip-Flop Based on Charge Sharing for Timing Slack Monitoring

POCHANA ASHOK KUMAR , MTECH(ES&VLSI) , MRCE

***Abstract*— Timing error predictors have a strong potential to reduce the worst case timing margins by monitoring timing slack of a design. However, these timing error predictors incur substantial amount of silicon area and power which limit the overall benefits in the system level. This paper presents a low overhead warning flip-flop (FF), which predicts setup time violations. It consists of a delay buffer and a warning generator along with a conventional master-slave FF. Low overhead FF can be designed by exploiting the concept of charge sharing to implement the warning generator. As the warning generator requires only seven transistors to predict the timing violation, the proposed warning FF occupies 30% less area and consumes27% less power compared to the state-of-the-art timing error predictors. A test chip is fabricated using the proposed FF in a 130-nm CMOS technology to verify the functionality of the proposed warning FF in dynamic voltage and frequency scaling applications. Measurement results from the test chip show that a performance improvement of 44% can be achieved at a supply voltage of 0.9 V by employing the proposed technique compared to the worst case design. For a typical chip, the power consumption can be reduced by 36% compared to the worst case design.**

***Index Terms*— Charge sharing, dynamic frequency scaling (DFS), dynamic voltage scaling (DVS), setup time violation, timing error predictors, timing margins.**

I. Introduction

PROCESS, voltage, and temperature (PVT) variations in scaled technology nodes cause significant performance uncertainty in the digital designs. Timing or supply voltage guard bands are added to maximum operating frequency (MOF) or minimum supply voltage to cope with PVT variations. However, these guard bands severely limit the performance and/or increase the power consumption of a design in the typical or best conditions. Moreover, transistor aging degrades the performance of a design with time. Hence, guard bands need to be added considering the life time of the design. As a result, traditional worst case design methodology is not suitable to implement energy-efficient designs as large guard bands are required in nanometer technology nodes. This motivates to implement design methodologies which can reduce the guard bands.

Traditionally, process monitors have been proposed to monitor the manufacturing process condition of a chip. In these techniques, body biasing is used to change the threshold voltage of the transistors based on process conditions. However, these techniques can address only global process variations but not local process and dynamic variations. Critical path replica circuits [1]–[4] have been explored to track the delay of the critical paths of a design. In this approach, a replica circuit which is strongly correlated with the actual critical path of the design is monitored to observe delay variations. This method can address global variations but not the local variations due to mismatch in the delay of the actual critical path and replica path. Moreover, the activation of critical path depends on the input data pattern.

On the other hand, *in situ* timing error monitoring techniques [5]–[8] can address both local and global variations by monitoring the timing slack. These techniques directly monitor the output of combinational logic using a specialized flip-flop (FF). An error signal is flagged in the case of timing violation. Hence, the supply voltage or frequency of design can be altered by monitoring the error signal. These techniques are mainly classified into two categories: error detectors [5], [6] and error predictors [8], [9]. Error detectors such as Razor I [6] and Razor II [7] detect the timing errors after their occurrence and correct the timing violations using architectural replay mechanism [7]. However, error detectors introduce a significant minimum path delay constraint which causes large area overhead because of buffer insertion. Moreover, architectural replay mechanism employed to correct timing violations is available in high-performance processors but not in application-specific integrated circuits (ASICs). Bubble Razor [10] does not incur area overhead for short path padding. However, this technique uses latches in the pipelines instead of FFs.

Error predictors [11], [12] flag a warning signal before the occurrence of timing violations by monitoring the delayed data. As the output of FF is always correct, these techniques do not incur correction overhead. Error predictors are suitable to implement ASICs as they do not require a correction mechanism. However, these techniques can monitor only gradual change in the delay of the critical paths. Canary FF [8], [9], [11] falls into this category. Canary FF employs a double-sampling architecture to predict the timing violations. Canary FF incurs large area and power overhead because of the shadow FF and delay buffers. Moreover, output of the shadow FF may enter into metastable state.

In [13], an aging sensor has been proposed, which uses delayed clock to create a guard band interval before the rising edge of the clock. As the guard band interval is created using the delayed edge of the previous clock cycle, the design is a function of operating frequency. An alternative version is also proposed in [13], which uses double-sampling architecture similar to canary FF. A timing monitoring circuit presented in [14] uses a sensor to predict the timing violations and a warning window generator to create a detection window. In this approach, the detection window is generated by using a transition detector in the clock tree which makes the clock tree implementation difficult.

A warning detection sequential in [12] and [15] observesthe delayed data transition during a warning interval to predict timing violations. This FF is not area and power efficient as a large number of delay buffers are required to delay the input data and to implement the edge detector. In [16], a sensor has been proposed which monitors the delayed master output to predict the timing violations. It requires a large area and power as it uses double-sampling architecture. A pre-error FF in [17] monitors data transition during the negative half cycle of the clock cycle. As warning margin is determined by negative half cycle of the clock, the design requires a clock of more than 50% duty cycle for better performance of the design.

The pulse-based error predictor in [18] monitors delayed master latch output in the high phase of the clock. However, edge detector used in this design requires more area and consumes more power. *In situ* monitors proposed in [19]–[21] observe master latch output and employ double-sampling architecture similar to the canary FF [8]. A timing error predictor in [22] employs double-sampling architecture similar to the canary FF except that a tunable delay buffer is used to delay the data signal. Double-sampling architectures require significant area and power overhead. A current-based timing error detector is proposed in [23]. A nine transistor transition detector [24] is designed to operate at supply voltage range 0.44–1.1 V for low-power applications. Critical path identification becomes difficult with increased variations in nanometer technology nodes [25]. Because of this, more number of critical paths have to be monitored using the timing error predictors to avoid functional failure. Hence, this paper presents a low area and power overhead timing error predictor for timing slack monitoring. In addition, the proposed warning FF can be used as an aging sensor similar to the work reported in [13] and [26].

The remainder of this paper is organized as follows.

The operation of the proposed warning FF is discussed in Section II. Chip implementation details are described in Section III. Measurement results are presented in Section IV. Section V discusses the comparison of timing error predictors and simulation results. Finally, this paper is concluded in Section VI.

II. Proposed Charge Sharing-Based Warning Flip-Flop

Warning FF is used to predict the setup time violation. Warning FF flags a warning signal, if data transitions at the input of FF happen during a timing window before the rising edge of the clock. Traditionally, either the input data of FF [11], [15] or master latch output [18], [20] is monitored to predict the setup time violations. In both approaches, the concept of delayed data is employed to predict the timing violations. Moreover, these approaches use either double- sampling technique or transition detection technique to flag warning signal. Based on these, warning FFs can be classified into four categories:

1) samples at input data *D* of FF and uses double-sampling technique [11];
2) samples master latch output *Mout* and uses double-sampling technique [20];
3) samples at input data *D* of FF and uses transition detection technique [15];
4) samples master latch output *Mout* and uses transition detection technique [18].

If the input data of the FF is monitored to flag warning signal, the delay of buffer should account for both setup time of FF and warning margin. However, if the master latch output is monitored to flag warning signal, the delay of buffer should account for warning margin only as the master latch delay already accounts for the setup time of FF.

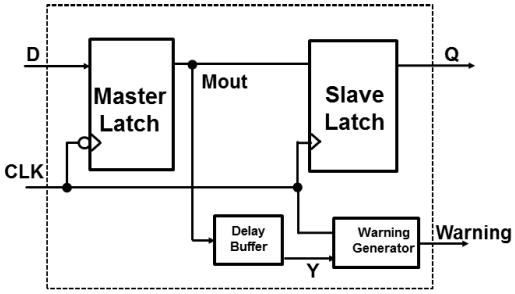


Fig. 1. Proposed warning FF.

The proposed warning FF is based on monitoring the delayed data at the output of master latch. It consists of a delay buffer and a warning generator along with a conventional master-slave FF to predict the setup time violations. The delay of the buffer is the warning margin of the proposed FF. If a data transition happens during a timing interval equal to the warning margin before the setup time window of the master latch, the proposed FF flags the warning signal. The warning signal is an active low signal. In the case of early data arrival at the input of FF, the warning generator output is logic high. However, in the case of late data arrival, the warning generator output is logic low. The schematic of the proposed warning FF is shown in Fig. 1.

A. Operation of the Proposed Flip-Flop

A data transition at the input of the FF can happen in either of the four timing windows shown in Fig. 2. A data transition during the window 1 is treated as an early data arrival. In this case, the available timing slack is very high. A data transition during window 2 is treated as late data arrival. In this case, the available timing slack is less. A data transition in this window means that a timing violation may occur if the delay of the critical path is further increased due to PVT variations. A data transition in timing window 3 leads to setup violation,

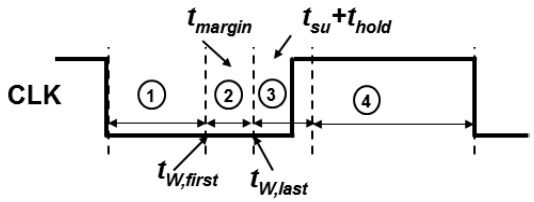


Fig. 2. Timing windows for possible data transition.

TABLE I
OPERATION OF THE PROPOSED WARNING FF

Timing Window	Data Transition	Before the rising edge of clock			After the rising edge of clock			Q	Warning
		D	Mout	Y	D	Mout	Y		
1	Rise	1	1	1	X*	1	1	1	1
	Fall	0	0	0	X	0	0	0	1
2	Rise	1	1	0	X	1	1	1	0
	Fall	0	0	1	X	0	0	0	0
3	Rise or Fall	X	PS**	PS	X	PS/D	PS/D	PS/D	1/0
4	Rise or Fall	X	PS	PS	X	PS	PS	PS	1

* Don't care ** Previous State

and the output of master latch may enter into metastable state. So, the maximum delay constraint of the design should satisfy (1), so that the data transitions under PVT variations can happen before the timing window 3. The maximum delay constraint is given by

$$T_{\max} \leq T_{\mathrm{CLK}} - t_{\mathrm{ClktoQ}} - t_{\mathrm{su}} - t_{\mathrm{margin}} \quad (1)$$

where $T_{\max}$ is the maximum combinational path delay, T_{CLK} is the clock period, t_{ClktoQ} is the previous stage FF clock to Q delay, t_{su} is setup time, and t_{margin} is warning margin. A data transition during the timing window 4 does not appear at the output of master latch as the master latch is opaque in this window.

The operation of the warning FF is given in Table I. If a data transition occurs during the timing window 1, the master latch output (*Mout*) and delayed master latch output (*Y*) (refer to Fig. 1) are the same in both low and high phases of the clock. However, if a data transition happens in the timing window 2, the delayed master output (*Y*) is different from *Mout* in low and high phases of the clock. In this case, the warning signal becomes low. However, in both the cases, the FF samples the correct value. When a data transition happens during the timing window 3, the master latch output (*Mout*) may enter into metastable state and it can resolve to either logic "0" or logic "1". Hence, the delayed master output (*Y*) may store the present data (*D*) or the previous state of FF. The conceptual timing diagram of the proposed warning FF is shown in Fig. 3.

In clock cycle I, the data transition happens early, before the warning margin. The master latch samples the correct value. The delayed master latch output (*Y*) makes a transition during the negative half cycle of the clock represented as *A*1. In this case, delayed master output (*Y*) is the same in both the low and high phases of the clock. Hence, the warning generator output is logic high signal. The delay between the input of FF (*D*) and the master latch output (*Mout*) is the setup time of master latch. In clock cycle II, data transition

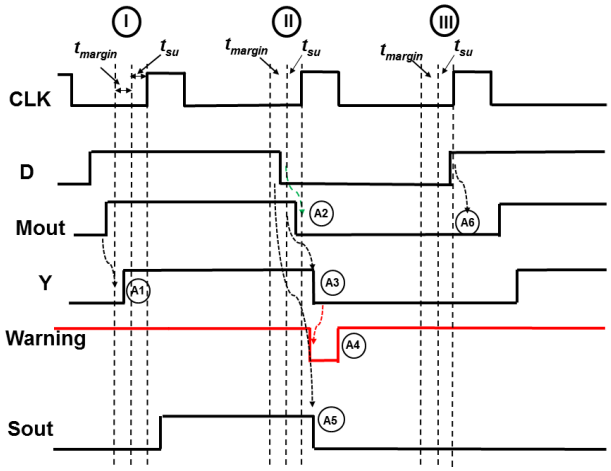


Fig. 3. Conceptual timing diagram of the proposed warning FF.

happens during the warning margin interval, the master latch samples the valid data represented as *A*2. However, the delayed master latch output makes transition during the high phase of the clock represented as *A*3. In this case, the delayed master output is different in high and low phases of the clock. So, the warning generator output becomes low represented as *A*4. Even the warning signal is flagged, the output of FF is always correct represented as *A*5. So, there is no correction overhead with this approach. In clock cycle III, the data transition happens during the setup time window of the master latch. This data transition does not appear at the output of master latch represented as *A*6. Hence, the proposed FF does not detect data transitions happened during this window. Similarly, data transitions during the high phase of clock will not appear at the master latch output as it is opaque during the high phase of the clock. Hence, the proposed FF does not suffer from short path problem. The minimum delay constraint of the design is given by

$$T_{\min} \geq t_{\mathrm{hold}} - t_{\mathrm{ClktoQ}} \quad (2)$$

where $T_{\min}$ is the minimum combinational path delay, t_{ClktoQ} is the minimum clock-to-Q delay of the previous stage FF, and t_{hold} is the hold time of FF.

The detectable slack of the proposed warning FF is defined as the difference between the point-of-first-warning (POFW) ($t_{W,\mathrm{first}}$) and the point-of-last-warning ($t_{W,\mathrm{last}}$) (refer to Fig. 2). Ideally, the detectable slack is the delay of buffer, which is used to delay the master latch output. Hence, the detectable slack is given by

$$t_{d,\mathrm{slack}} = t_{d,\mathrm{buffer}} \quad (3)$$

where $t_{d,\mathrm{slack}}$ is the detectable slack and $t_{d,\mathrm{buffer}}$ is the delay of the buffer.

B. Warning Generator

The warning generator is designed using the concept of charge sharing. It requires only seven transistors to detect a transition during the high phase of the clock. The schematic of the warning generator is shown in Fig. 4. The delayed master output drives two cascaded-clocked inverters. During the low

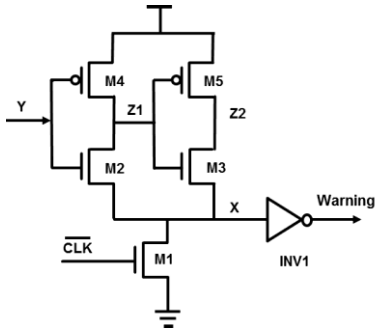


Fig. 4. Schematic of the warning generator.

TABLE II
OPERATION OF WARNING GENERATOR

Timing Window	Data Transition	CLK=0				CLK=1				Warning
		Node Y	Node Z1	Node Z2	Node X	Node Y	Node Z1	Node Z2	Node X	
2	Rise	1	0	1	0	1	0	1	0	1
	Fall	0	1	0	0	0	1	0	0	1
3	Rise	0	1	0	0	1	Dischraged	Charged	Charged	0
	Fall	1	0	1	0	0	Charged	Dischraged	Charged	0
4	Rise	0	1	0	0	0	1	0	0	1
	Fall	1	0	1	0	1	0	1	0	1
	Rise	0	1	0	0	0	1	0	0	1
	Fall	1	0	1	0	1	0	1	0	1

phase of the clock, these inverters behave like conventional inverters and the intermediate node *X* is discharged to logic zero. However, during the high phase of the clock, node X becomes floating as the transistor M1 is OFF. The warning signal is at logic high state. Whenever the input of warning generator is different in low and high phases of the clock, the intermediate node *X* charges from the nodes *Z*1 and *Z*2 for rise and fall data transitions, respectively. However, the charged voltage at intermediate node *X* is less than supply voltage for rise and fall transition at node *Y*. The inverter INV1 should treat the charged voltage at node *X* as logic high-input voltage. For this, the switching threshold of inverter should be less than charged voltage at node *X*. The inverter INV1 is skewed to change the switching threshold. The switching threshold of the inverter is less than half of the supply voltage.

The operation of the warning generator is given in Table II. The delayed master output (*Y*) depends on timing window (refer to Fig. 2) in which input data (*D*) transition happens. An input data (*D*) transition may lead to: 1) a transition at *Y* during the low phase of the clock; 2) a transition at *Y* during the high phase of the clock; and 3) no transition until the next low phase of the clock. A transition at *Y* during the low phase of the clock occurs, if the data (*D*) transition happens in timing window 1. A transition at *Y* during the high phase of the clock occurs, if the data (*D*) transition happens in timing window 2. For a data transition in timing windows 3 and 4, no data transition happens at *Y* until the next low phase of the clock. The detailed operation of the warning generator is discussed in Sections II-B.1–II-B.3.

1) Data (Y) Transition During Clock Low Phase: If a signal transition at *Y* occurs during the low phase of the clock, the state of the node *Y* is the same in both low and high phases of the clock. So, the intermediate nodes *X*, *Z*1, and *Z*2 do not get affected after the rising edge of the clock. When CLK = 0, the intermediate node *X* is discharged. Hence, the warning signal is high.

2) Rise Transition (Y) During Clock High Phase: For a rise transition at the input of the warning generator (*Y*) in the high phase of clock, the warning generator input is initially logic zero. A logic zero input during the low phase of clock, charges the intermediate node *Z*1 and discharges the intermediate node *Z*2. The initial states of nodes *X*, *Z*1, and *Z*2 are shown in Fig. 5(a). The transistor M1 is ON, during the low phase of the clock. So, the node *X* is discharged to logic zero. In addition, the nodes *Z*1 and *Z*2 are logic high and low, respectively. After a rise transition at the input of warning generator in the positive half cycle of clock, the states of nodes *X*, *Z*1, and *Z*2 are shown in Fig. 5(b). In this case, the transistor M1 is *off*. So, the intermediate node *X* is charged through the transistor M2. Because of charge sharing between nodes *X* and *Z*1, the voltage at node *X* starts increasing toward logic high. If this voltage is more than the switching threshold of the inverter, the warning signal becomes low. In this way, a rise transition at delayed master output is detected.

3) Fall Transition (Y) During Clock High Phase: For a fall transition at *Y* in the high phase of the clock, the delayed master output should be logic high in the negative half cycle of the clock. The initial states of nodes *X*, *Z*1, and *Z*2 are shown in Fig. 5(c). The transistor M1 is ON during the low phase of the clock, so the warning signal is high during this interval. The intermediate nodes *Z*1 and *Z*2 are logic low and high, respectively. After a fall transition in the high phase of the clock, the node *X* is charged from node *Z*2 through transistor M3. This charge sharing leads to a rise transition at node *X*. The output of warning generator becomes low. The states of nodes *X*, *Z*1, and *Z*2 in the high phase of the clock are shown in Fig. 5(d).

III. CHIP IMPLEMENTATION DETAILS

A two-stage pipelined design is implemented to verify the functionality of the proposed warning FF. The internal blocks of the test chip is shown in Fig. 6. The implemented blocks are divided into three sections: 1) input section; 2) pipelined design; and 3) output section. The input section consists of a serial-in parallel-out (SIPO) shift register and a toggle circuit. The pipelined design section consists of two-stage pipelined circuit with the proposed warning FF and an AND tree to group the warning signals. The output section consists of a parallel-in serial-out (PISO) shift register.

A 13-bit SIPO shift register is used in the implementation of the design. Out of 13 bits, 8 bits for input data of the pipelined design, 1 bit for enable signal of the toggle circuit, 2 bits for selecting the critical path delay, and 2 bits for selecting the warning margin of the warning FF are assigned. To check the functionality of the proposed warning FF on chip, the critical path needs to be activated in each clock cycle, which can be done by changing the inputs of the pipelined design using a toggle circuit. The data input to the design is given from an SIPO shift register through the toggle circuit. Toggle circuit has an enable signal. If the enable signal is high, the inputs of the design will change in every clock cycle which is required to evaluate the warning FF in the pipelined design. In this case, the input data pattern stored in the SIPO register can be inverted in each clock cycle using the toggle circuit and

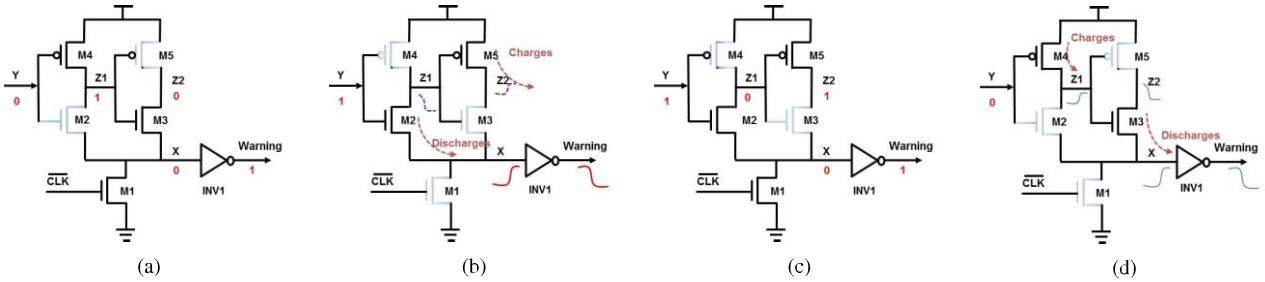


Fig. 5. Operation of warning generator for data transitions in timing window 2. (a) Before rise transition. (b) After rise transition. (c) Before fall transition. (d) After fall transition.

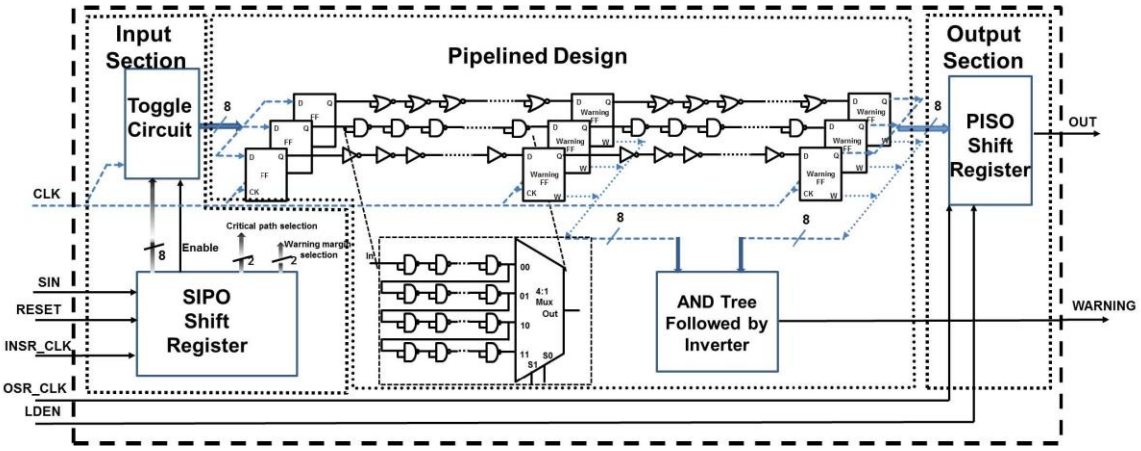


Fig. 6. Details of chip implementation.

is provided to the pipelined design. As all the paths in the pipelined design are made critical at the design time, all the paths have 100% activity during run time.

Each stage of the pipelined design has eight paths which are designed using inverter, NAND and NOR chains. The proposed warning FF is inserted at the end points of all the critical paths. The length of the critical path can be modified inside the chip using a configurable delay chains with a 4:1 multiplexer in the critical path. Hence, this scheme allows operating the same design at different clock frequencies. The warning signal from all the warning FFs are grouped into a single warning signal using an AND tree. An inverter converts the AND-tree output to an active high warning signal WARNING as the warning signal in the FF level is active low. The delay of the AND tree should be less than the minimum pulsewidth of the warning signal of the proposed FF and is given by

$$t_{d,\text{AND-tree}} = T_{\text{PW,min}} \quad (4)$$

where $t_{d,\text{AND-tree}}$ is the delay of AND tree and $T_{\text{PW,min}}$ is the minimum pulsewidth of the warning signal of the proposed FF.

An 8-bit PISO shift register is used to load the outputs of the pipelined design. The PISO shift register has two external inputs: one is the clock (OSR_CLK) to the shift register and the other is a control signal (*LDEN*) to enable the shift register to load from the pipelined design and to shift the data of the shift register. If the LDEN is high, the outputs of the pipelined design are loaded into PISO shift register in every clock cycle.

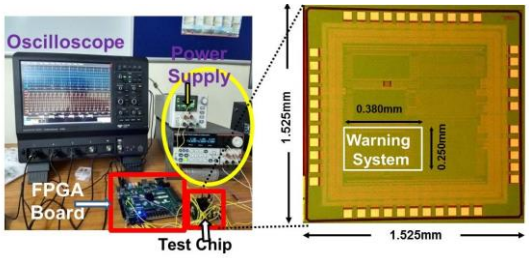


Fig. 7. Experimental setup.

If the LDEN is low, the stored data of PISO shift register is serially shifted out in every clock cycle.

IV. Measurement Results

The experimental setup to measure the test chip is shown in Fig. 7 which shows a test chip implemented in 130-nm CMOS technology, mounted on QFN-48 socket, a field-programmable gate array (FPGA) board and an oscil- loscope. The SIPO shift register inside the chip is programed through the FPGA Board (ZedBoard) to provide the input data. The serial output of the PISO register is monitored continuously. In the case of warning, all the outputs of the design are verified by shifting the PISO register.

The frequency of baseline or worst case design with 10% voltage drop, a temperature of 85 °C, and 2-sigma process variations is obtained by using the method of [22].

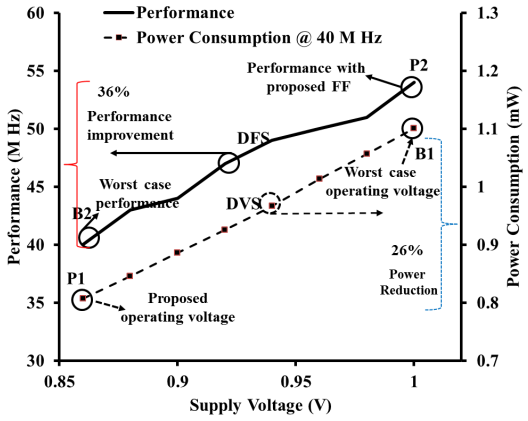


Fig. 8. Performance and power consumption of the design.

The frequency of the baseline design is 40 MHz for a supply voltage of 1 V. The conventional operating voltage and baseline frequency of the design are represented as *B*1 and *B*2, respectively, in Fig. 8. However, by employing the proposed FF in the pipelined design with dynamic voltage scaling (DVS), the pipelined design can operate at a reduced voltage represented as *P*1 for the same performance which results in 26% savings in the power consumption. For dynamic frequency scaling (DFS), the supply voltage is kept constant and frequency of the design is increased. The pipelined design with the proposed FF and DFS can operate at an increased frequency represented as *P*2 which leads to 36% performance improvement at a supply voltage of 1 V. This shows the advantage of using the proposed FF in reducing the timing or voltage guard bands.

A. Impact of Supply Voltage Scaling

Conventionally, the design will operate at a supply voltage of 1.2 V considering worst case margins. However, the pro- posed warning FF can detect the actual operating condition which in turn can reduce the worst case guard bands. Initially, the designs are operated at a supply voltage of 1.2 V with an operating frequency of 50 MHz. The supply voltage of the design is reduced until the warning signal is asserted. The supply voltage for which warning signal becomes high is defined as POFW voltage. If the supply voltage of the design is further reduced, the warning signal is still high and a functional error may occur in the design. A system failure is verified by monitoring the outputs of the design using PISO shift register. The voltage for which a functional failure occurs is known as point-of-first-failure (POFF) voltage. The differ- ence between the POFF and POFW voltages is the available margin for warning. This warning margin is a function of number of delay buffers used to delay the master latch output of the proposed warning FF.

The POFW voltage is measured for 26 chips by reducingthe supply voltage manually in steps of 1 mV until the warning signal becomes high. As the implemented test chip does not include an adaptive controller, the supply voltage is reduced manually while monitoring the warning signal.

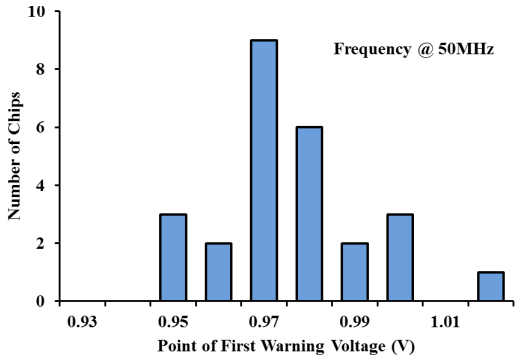


Fig. 9. POFW voltage of measured chips.

However, an external controller used in [15] or on-chip controller proposed in [22] can be employed to change the voltage or frequency of the design. Fig. 9 shows histogram of the POFW voltage for 26 measured chips operating at 50 MHz and at a temperature of 25 °C. It is evident from the POFW voltage of measured chips that the design can operate at reduced supply voltage for a given performance compared to a worst case design. For a typical chip among the measured chips, the supply voltage can be reduced to 0.97 from 1.2 V. However, the minimum voltage required for the worst chip among the measured chips is 1.015 V. Among 26 chips,14 chips can operate at a supply voltage less than 0.97 V for the same performance. Moreover, the best chip among the measured chips requires a supply voltage of 0.94 V, which is 75 mV less than that of the worst chip. This reduction in sup- ply voltage results in lower power consumption, and it shows the significance of warning FF for low-power applications.

To measure the dynamic power consumption of the design (P_{Design}), the following steps are used. First, the power consumption of the design along with input-output (IO) cells (P_{D_IO}) is measured. Then, the power consumption of only IO cells (P_{IO}) is measured by not applying the clock to design. However, three IO cells (clock, output, and warning signals of design) are active, only when the design is working. So, the power consumed by IO cells (P_{IO}) does not include the dynamic power of these three IO cells. So SPICE simulations are carried out to find the dynamic power of these IO cells ($P_{\text{SIM\_IO}}$). To increase the accuracy in the measurement of power con- sumption, the dynamic power of three IO cells ($P_{\text{SIM\_IO}}$) is subtracted. The dynamic power consumption of the design is given by

$$P_{\text{Design}} = P_{D\_IO} - P_{\text{IO}} - P_{\text{SIM\_IO}}. \quad (5)$$

The abovementioned method is used to measure the power consumption in this paper as the same supply voltage is connected for IO cells and the proposed design. The implemented design is having a dedicated clock signal. The histogram of dynamic power consumption of the design for 26 measured chips is shown in Fig. 10. The chips are operated at the POFW to measure the power consumption of the design. The maximum power consumed by the design is 1.35 mW, whereas the minimum power consumed by the design is 1.20 mW.

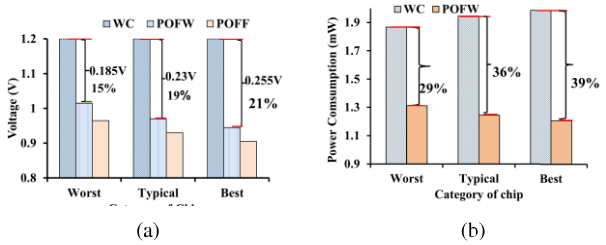


Fig. 11. Comparison of worst, typical, and best chips among the 26 measured chips. (a) POFW voltage. (b) Power consumption.

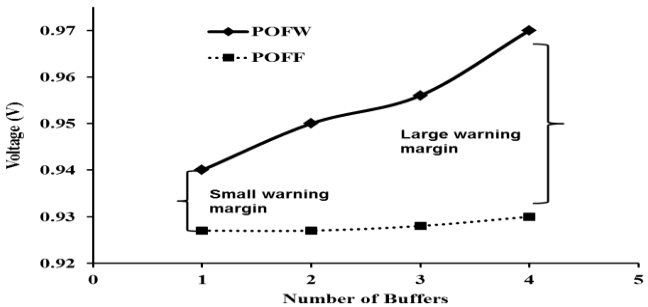
The power consumption of a typical chip among the measured chips is 1.26 mW. Among 26 measured chips, 15 chips can work with a power consumption less than 1.26 mW.

The POFW and POFF voltages for three different categories of chips are shown in Fig. 11(a). The POFW voltage of the worst and typical chips, out of measured chips is 185 and230 mV less than the nominal supply of 1.2 V, respectively. The power consumption of the proposed design with DVS is measured by operating the design at a POFW voltage. For a typical chip, the power consumption is reduced by 36%, whereas for the best chip, the power consumption is reduced by 39% using the proposed warning FF with DVS. The power consumption of three different categories of chips is shown in Fig. 11(b).

B. Impact of Delay Buffers on POFW Voltage

In the proposed approach, a tunable delay buffer used in [15] is employed to change the warning margin of the proposed FF. This tunable delay buffer can be used to provide different warning margins for varying process conditions. A process monitor can be used to detect global process conditions. The warning margin of the proposed FF is varied using the select lines of the multiplexer. In this case, a 4:1 multiplexer is employed for selecting the warning margin.

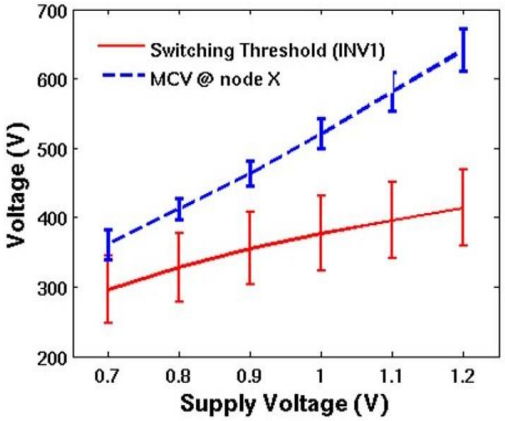
For a select input of 00, the critical path delay is minimum and POFW occurs at a voltage of 0.94 V for a typical chip. In this case, maximum power savings can be achieved. However, the design can tolerate less delay variations as the warning margin is small. For a select input of 11, the warning margin is maximum and the POFW occurs at a voltage of 0.97 V for a typical chip. However, in this case, the design can tolerate large delay variations as the warning margin is maximum. In all the cases, the POFF voltage is the same, as it is determined by the setup time of the FF. Fig. 12 shows the POFW and POFF voltages for a typical chip as a function of delay buffers.



C. Impact of Frequency Scaling

The frequency of baseline or worst case design at a supply voltage can be obtained by using the method of [22]. First, the MOF of each chip is measured at room temperature with 10% drop in the supply voltage. Second, Gaussian fitting of measured MOF of all chips is used to find the mean (μ) and standard deviation (σ). Then,2-sigma frequency (2σ) and 5% of the mean value are reduced from the mean value to account for process variations and temperature margin of 60 °C, respectively. As the design is measured at room temperature, the timing margin to account for the temperature of 85 °C is found by SPICE simulations. It is found from the simulation results that the performance of the design at a temperature of 85 °C is degraded by 6.4% and 5% at a supply voltage of 1.05 and 0.9 V, respectively, compared to performance at room temperature (25 °C). For an operating voltage of 1 V, the MOF of the measured chips at room temperature with 10% voltage drop is shown in Fig. 13. The mean and standard deviation are found to be 45 and1.7 MHz, respectively. The baseline frequency is 40 MHz atan operating voltage of 1 V. Similarly, the baseline frequencies for supply voltage of 0.9, 0.95, and 1.05 V are 31, 35, and44 MHz, respectively.

The MOF of three categories of the chips (worst, typical, and best) is shown in Fig. 14. Out of 26 measured chips, the chip which flags a warning signal for the lowest frequency at a fixed voltage is treated as the worst chip. Similarly, the chip which flags a warning signal for the highest frequency at a fixed voltage is treated as the best chip. The best chip can perform 23% better than a worst chip at a supply voltage of 0.9 V. The same chip can perform 14% better than a worst chip at a supply voltage of 1.05 V. This shows the advantage of using warning FF at lower supply voltage. The proposed warning FF can improve the performance of the design by 44% and 31% compared to baseline design at a supply voltage of 0.9 and 1.05 V, respectively, under typical operating conditions.



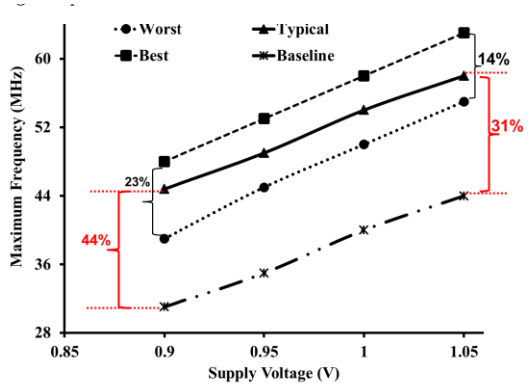


Fig. 14. Performance improvement of worst, typical, and best chips at room temperature.

V. Simulation Results

Section IV discussed the advantages of using the proposed warning FF. The implementation details of warning FF, such as the maximum charged voltage (MCV) at the intermediate node, and sizing of the transistors in the warning generator are discussed in this section. Moreover, the FF-level comparison is discussed to quantify area and power savings with respect to the warning FFs available in the literature. The proposed FF is implemented in industrial 130-nm CMOS technology. Postlayout simulations on parasitic extracted netlist have been carried out using HSPICE.

A. Warning Generator

The warning generator (refer to Fig. 4) is designed using the concept of charge sharing. The MCV at node *X* of the warning generator and switching threshold of the skewed inverter are critical factors for the proper functionality of warning FF. To verify the MCV at node *X* for data transitions at the input of warning generator, 10 000 Monte Carlo simulations have been performed with 3-sigma process variation for a supply voltage range 0.7–1.2 V at a temperature of 25 °C. The minimum value of MCV is used to determine the switching threshold of the skewed inverter INV1 (refer to Fig. 4). The MCV at node *X* and switching threshold of the skewed inverter (INV1) are shown in Fig. 15.

Postlayout simulations on parasitic extracted netlist have been carried out to verify the operation of the proposed

Fig. 16. Timing diagram of the warning generator.

warning generator. Fig. 16 shows the timing diagram of CLKB (inverted clock), input of warning generator (*Y*), warning signal, intermediate nodes *X*, *Z*1, and *Z*2 for rise and fall transitions. For both rise or fall transitions at node *Y*, when CLKB = 0, the intermediate node *X* is charged to voltage greater than half of the supply voltage. For a rise transition, the intermediate node *Z*1 is discharged to intermediate value, which charges the node *X* and the node *Z*2 is charged to supply voltage. Similarly, in the case of fall transition, node *Z*2 is discharged to an intermediate value which charges the node *X* and node *Z*1 is charged to supply voltage. In both cases, the warning signal is active low pulse. The delay of the warning generator is increased by 5× at a supply voltage of 0.7 V compared to the delay of warning generator at 1.2 V. The variability of the delay of warning generator is 8.7% and
32% at supply voltages of 1.2 and 0.7 V, respectively, with
3-sigma process variations (including both within-die variation and die-to-die variation).

B. Flip-Flop-Level Comparison

To quantify the area and power savings of the proposed warning FF compared to existing FFs, the canary FF [11], warning detection sequential [15], *in situ* monitor [20], and pulse-based timing error predictor [18] are implemented in industrial 130-nm CMOS technology. The comparison of different warning FFs is given in Table III.

1) Area: The proposed warning FF requires 1.95× area compared to the conventional FF. However, the proposed FF

TABLE III
COMPARISON OF WARNING FFs

Parameter	Proposed	Ref [11]	Ref [15]	Ref [20]	Ref [18]	DFF
Area (layout)	1.95×	3.45×	3.1×	3×	2.81×	1×
Number of buffers	1	2	3	1	2	NA *
Monitoring node	Master latch output	FF input	FF input	Master latch output	Master latch output	NA
Architecture	Charge sharing	Double sampling	Transition detection	Double sampling	Transition detection	NA
Clock to Q delay	1.05×	1.05×	1.01×	1.03×	0.97×	1×
Setup time(ps)	113.6	71	77	117.8	187.76	76
Hold time(ps)	-10.3	1.9	1.9	-13.2	-17	2
Clock power (μW)	4.62	8.51	4.05	8.49	4.04	3.97
Average Power (No warning)(μW)	11.5	20.45	15.6	19.2	15.8	7.32
Average Power (Warning)(μW)	14.5	19.21	17.1	19.2	17.3	NA
Peak Power (No warning)	1.14×	2.45×	1.12×	2.42×	1.82×	1×
Peak Power (Warning)	1×	1.66×	1.52×	1.81×	1.33×	NA

* Not Applicable

occupies 30% less area compared to [18]. This area savings are due to the requirement of less transistors to implement the warning generator. Moreover, the approach in [18] requires two delay buffers, one for warning margin and the other for transition detection unlike the proposed warning FF which requires only one delay buffer to provide warning margin. The area of canary FF [11], warning detection sequential [15], *in situ* monitors [20] and pulse-based predictor [18] is 3.45×,
3.1×, 3×, and 2.81×, respectively, compared to that of conventional FF as given in the second row of Table III. The large area overhead in [11] and [20] is due to the usage of double-sampling architecture. Moreover, the canary FF [11] requires extra delay buffers to account for the setup time of the main FF as it monitors at the input of FF.

2) Average Power: The proposed warning FF consumes 1.57× more power compared to the conventional FF. However, the proposed FF consumes 27% less power compared to the pulse-based timing error predictor [18] with 50% data activity at an operating frequency of 250 MHz. The reduction in power consumption is due to the less number of the transistors in the warning generator. Moreover, clock power of the proposed warning FF is just 1.16× to that of the conventional FF with 0% data activity. However, Canary FF [11] and *in situ* monitor [20] consume 2.14× and 2.13× more power, respectively, compared to the conventional FF with 0% data activity. This increased power consumption in [11] and [20] is due to the usage of more number of clocked transistors in the double-sampling architectures. In the proposed warning FF, the power consumption is more with the warning signal asserted compared to no warning condition. However, Canary FF consumes less power with warning signal asserted because of the reduced data activity at shadow FF compared to no warning condition. The power consumption due to the warning window generator of [15] is not taken into account for the comparison as it can be shared among warning FFs.

3) Peak Power: The peak power of the proposed FF is measured for different data to clock durations. When the data transition happens during the warning margin window, the peak power is higher compared to peak power due to a data transition before the warning margin window. The peak power of the different FFs is given in Table III. The double-sampling techniques [11], [20] consume higher peak power compared to transition detection approaches [15], [18]. The proposed FF consumes minimum peak power, out of the compared FFs, when the warning signal is flagged. When there is no warning signal, the proposed FF consumes 14% extra peak power compared to the conventional DFF. The warning detection sequential of [15] consumes 12% extra power compared to the conventional DFF.

TABLE IV
POWER SAVINGS OF ISCAS89 BENCHMARK CIRCUITS

Benchmark Circuit	Frequency (MHz)	%Replacement Rate	%Area Overhead	POFW (V)	%Power Savings
S13207	416	7	7.2	0.94	18
S15850	250	11.2	5.5	1	19
S38417	225	13.23	6.5	1	22

C. ISCAS89 Benchmark Circuits

The ISCAS89 benchmark circuits [27] are implemented to operate under worst case conditions with slow process corner, 125 °C temperature, and 10% drop in the supply voltage. The critical paths with slack less than 20% of time period are monitored with proposed FFs. The replacement rate is the ratio of number of proposed warning FFs to the total number of FFs in the design. Initially, the designs are simulated using a gate-level simulator (Synopsys VCS) under typical operating conditions with a nominal supply voltage of 1.2 V at a temperature of 25 °C. The timing libraries for different supply voltages are created using Cadence Liberate. The supply voltage of the design is reduced in steps of 20 mV until a warning signal is flagged. The POFW voltage is found by simulations for each design. From the simulation results, it is evident that by using the proposed FF, the designs can operate at reduced voltage under typical conditions. Power savings up to 22% can be obtained by employing the proposed FF in ISCAS89 benchmark circuits. The power savings obtained using the proposed FF in ISCAS89 benchmark circuits are given in Table IV.

VI. Conclusion

DVS and DFS with timing error monitors have become an effective way to reduce the worst case timing guard bands. This paper presented a low overhead warning FF, which monitors delayed master latch output to predict the timing vio- lations. The proposed warning FF consumes 27% less power at 50% activity factor and requires 30% less area compared to the timing error predictors available in the literature. A test chip is fabricated in industrial 130-nm CMOS technology to verify the effectiveness of the proposed warning FF in reducing the timing margin. Measurement of the test chip demonstrates that a design with the proposed warning FF can operate at
44% and 31% higher clock frequency at a supply voltage of 0.9 and 1.05 V, respectively, compared to the traditional worst case design in typical conditions. For a typical chip, the power consumption can be reduced by 36% compared to the conventional worst case design.

Acknowledgment

The authors would like to thank Europractice IC service for chip fabrication.

References

[1] T. Kuroda *et al.*, "Variable supply-voltage scheme for low-power high-speed CMOS digital design," *IEEE J. Solid-State Circuits*, vol. 33, no. 3, pp. 454–462, Mar. 1998.

[2] J. W. Tschanz *et al.*, "Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage," *IEEE J. Solid-State Circuits*, vol. 37, no. 11, pp. 1396–1402, Nov. 2002.

[3] J. T. Kao, M. Miyazaki, and A. P. Chandrakasan, "A 175-mV multiply-accumulate unit using an adaptive supply voltage and body bias architecture," *IEEE J. Solid-State Circuits*, vol. 37, no. 11, pp. 1545–1554, Nov. 2002.

[4] A. Drake *et al.*, "A distributed critical-path timing monitor for a 65 nm high-performance microprocessor," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2007, pp. 398–399.

[5] D. Ernst *et al.*, "Razor: A low-power pipeline based on circuit-level timing speculation," in *Proc. Int. Symp. Microarchitecture*, Dec. 2003, pp. 7–18.

[6] S. Das *et al.*, "A self-tuning DVS processor using delay-error detection and correction," *IEEE J. Solid-State Circuits*, vol. 41, no. 4, pp. 792–804, Apr. 2006.

[7] S. Das *et al.*, "RazorII: *In Situ* error detection and correction for PVT and SER tolerance," *IEEE J. Solid-State Circuits*, vol. 44, no. 1, pp. 32–48, Jan. 2009.

[8] T. Sato and Y. Kunitake, "A simple flip-flop circuit for typical-case designs for DFM," in *Proc. Int. Symp. Quality Electron. Des. (ISQED)*, Mar. 2007, pp. 539–544.

[9] H. Fuketa, M. Hashimoto, Y. Mitsuyama, and T. Onoye, "Adaptive performance compensation with in-situ timing error prediction for subthreshold circuits," in *Proc. Custom Integr. Circuits Conf. (CICC)*, 2009, pp. 215–218.

[10] M. Fojtik *et al.*, "Bubble Razor: Eliminating timing margins in an ARM Cortex-M3 processor in 45 nm CMOS using architecturally independent error detection and correction," *IEEE J. Solid-State Circuits*, vol. 48, no. 1, pp. 66–81, Jan. 2013.

[11] H. Fuketa, M. Hashimoto, Y. Mitsuyama, and T. Onoye, "Adaptive performance compensation with in-situ timing error predictive sensors for subthreshold circuits," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 20, no. 2, pp. 333–343, Feb. 2012.

[12] B. P. Das and H. Onodera, "Warning prediction sequential for transient error prevention," in *Proc. IEEE Int. Symp. DFT VLSI Syst.*, Oct. 2010, pp. 382–390.

[13] M. Agarwal, B. C. Paul, M. Zhang, and S. Mitra, "Circuit failure prediction and its application to transistor aging," in *Proc. VLSI Test Symp. (VTS)*, May 2007, pp. 277–286.

[14] B. Rebaud *et al.*, "Timing slack monitoring under process and envi- ronmental variations: Application to a DSP performance optimization," *Microelectron. J.*, vol. 42, no. 5, pp. 718–732, May 2011.

[15] B. P. Das and H. Onodera, "Frequency-independent warning detection sequential for dynamic voltage and frequency scaling in ASICs," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 22, no. 12, pp. 2535–2548, Dec. 2014.

[16] T. Azam and D. R. S. Cumming, "Efficient sensor for robust low-power design in nano-CMOS technologies," *IET Electron. Lett.*, vol. 46, no. 11, pp. 773–775, May 2010.

[17] M. Wirnshofer, L. Heiß, G. Georgakos, and D. Schmitt-Landsiedel, "A variation-aware adaptive voltage scaling technique based on in-situ delay monitoring," in *Proc. IEEE Int. Symp. Design Diagnostics Electron. Circuits Syst. (SDDECS)*, Apr. 2011, pp. 261–266.

[18] L.-Y. Chiou, C.-R. Huang, and M.-H. Wu, "A power-efficient pulse-based in-situ timing error predictor for PVT-variation sensitive circuits," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Jun. 2014, pp. 1215–1218.

[19] V. Huard *et al.*, "Adaptive wearout Management with *in-situ* aging monitors," in *Proc. IEEE Int. Rel. Phys. Symp.*, Jun. 2014, pp. 1–11.

[20] M. Saliva *et al.*, "Digital circuits reliability with *in-situ* monitors in 28 nm fully depleted SOI," in *Proc. Design, Autom. Test Eur. Conf. Exhib. (DATE)*, 2015, pp. 441–446.

[21] A. Benhassain *et al.*, "Timing in-situ monitors: Implementation strategy and applications results," in *Proc. Custom Integr. Circuits Conf. (CICC)*, 2015, pp. 1–4.

[22] W. Shan, X. Shang, L. Shi, W. Dai, and J. Yang, "Timing error prediction AVFS with detection window tuning for wide-operating-range ICs," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, to be published.

[23] Y. Zhang *et al.*, "iRazor: 3-transistor current-based error detection and correction in an ARM Cortex-R4 processor," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2016, pp. 160–162.

[24] X. Shang, W. Shan, L. Shi, X. Wan, and J. Yang, "A 0.44 V-1.1 V 9-transistor transition-detector and half-path error detection technique for low power applications," in *Proc. A-SSCC*, Nov. 2017, pp. 205–208.

[25] I. Kwon, S. Kim, D. Fick, M. Kim, Y.-P. Chen, and D. Sylvester, "Razor-lite: A light-weight register for error detection by observing virtual supply rails," *IEEE J. Solid-State Circuits*, vol. 49, no. 9, pp. 2054–2066, Sep. 2014.

[26] M. Omaña, D. Rossi, N. Bosio, and C. Metra, "Low cost NBTI degradation detection and masking approaches," *IEEE Trans. Comput.*, vol. 62, no. 3, pp. 496–509, Mar. 2013.

[27] (Nov. 2017). *ISCAS'89 Benchmarks*. [Online]. Available: http://www.pld.ttu.ee/maksim/benchmarks/iscas89/verilog/

Clock-Gating of streaming applications for energy efficient implementations on FPGAs

1. K. Mounika, M.Tech student, Mallareddy College Of Engineering

2. D. Rajendra Prasad, Assistant Professor, Mallareddy College Of Engineering

Abstract— **This project deals with reducing the dynamic power in streaming applications by using clock-gating. Streaming applications consists of broad class of computing algorithms used in areas such as cryptography, digital media coding ,communications, signal processing, video analytics , network routing etc.,. The power saving can be gained by selectively turning off or switching off parts of the circuit when they are temporarily inactive. So, that the switching state power can be reduced. In this the clock gating methodology is introduced in dataflow designs that are automatically included in the synthesis stage of the high level dataflow design flow. This concept describes an approach for developing energy optimized run-time reconfigurable design which is a benefit from clock gating. For this, FPGAs are highly desirable due to its reconfigurable (or) re-programmable nature, flexibility. It uses clock gating strategy for reducing dynamic power. The main aim of this project is to reduce dynamic power dissipation without much affecting performance and the experimental results also show that the applications synthesized on FPGA platforms shows that power reductions can be achieved with no loss in data throughput. By using this approach we can achieve low power , delay, area and low noise levels**.

Index Terms— **Clock-gating, data flow , dynamic power, high level synthesis**

I. INTRODUCTION

In general, power consumption is one of the major challenges in VLSI design performance. For a silicon device there are two components of power dissipation. They are static power dissipation and dynamic power dissipation. Static power dissipation is due to the leakage current produced within the transistor and by the ambient temperature. Dynamic

Dynamic power consumption can contribute upto 50% of the total power consumption. Inorder to reduce unwanted power dissipation additional circuit is added into the design that effectively clocks the gate. Power dissipation increases linearly with frequency due to highly influence of parasitic capacitances. To counteract this effect, ASIC designers have employed clock gating from past few years[1],[2],[3]. Clock gating is nothing but a power saving feature used in semiconductor devices which enables switching off or turning off circuits. Many silicon devices use clock gating to switch off controllers, parts of processor, bridges and buses inorder to reduces the power dissipation.

power dissipation is due to switching of transistors and by losses of charges being moved along wires.

Reducing power has also another benefits. They are it needs less stringent to cool the device, improved life of device or battery and low power costs. Due to these reasons, the power also frequently affects the choice of computing platform right at outset. For instance, Field-Programmable Gate Array (FPGA) imply higher power dissipation per logic unit by comparing with ASIC. This paper shows the impact of chosen technology on the architecture, but do not describe how to reduce the power at abstraction level of the design.

The major difference between FPGA and ASIC is ASIC is fixed implementation, that means these are pre-defined for a specific task, where as FPGA can be reprogrammed ON- site. FPGA eliminates non- recurring Engineering costs and thereby reduces time to market. Therefore FPGAs are highly desirable to implement digital systems due to their programmability , low end product life cycle, flexibility and all these makes ideal for small and medium applications. FPGAs are slower and less efficient due to added circuitry that is used to make flexible when compared to fixed implementation(ASIC).

Dynamic dataflow designs, for instance the RVC- CAL language possess interesting properties that are used for reducing the power without affecting the behavioral characteristics and the construction of the application. In RVC- CAL language, every actor can execute the processing tasks, executions may be disabled by the input blocking reads and communication between the actors can occur only by the order of storing lossless queue. As a result, an actor may be stopped for a period of time if its processing tasks are idle(or) output buffers (queues) are full without affecting overall throughput and semantical behavior of the design.

II. RELATED WORK & METHODOLOGY

In this paper, the work and methodology described below are based on Xilinx FPGA, which can be used to support the other architectures. A system which works or process on continues stream of bits is said to be streaming application. Cryptography, digital media coding, network routing, video analytics etc., are the examples. In clock gating scheme(CG),

the clock is given to those modules that are working at that instant. This clock-gating support the adding of additional logic to the existing synchronous circuit inorder to prune the clock tree, thereby disabling the portions of the circuit that are not in use. In an architecture, the additional device is called clocking circuit which is inserted before the data path unit which provides the clock inputs for working or active modules only. Hence it reduces dynamic power.

Related to our work, S.C Brunet, E. Bezati, C. Alberti M. Mattavelli, E. Amaldi, and J.W.Janneck proposed a multiple clock, domain-design methodology inorder to reduce the power consumption of dataflow programs. Their design motive was to optimize the mapping of the application along with meeting the desired design requirements. The optimization is gained by assigning optimized clock frequency to reduce power consumption.

Present FPGA support various clock-gating strategies and every manufacturer creates its own IP [5] for managing these approaches. The methodology explained here is based on primitives used specific to Xilinx FPGA architectures. Anyway these are modified inorder to support other FPGA vendor primitives.

The execution of a dataflow program contains a series of action firings [2],[6], which can be correlated to one another in a graph-based representation using an approach called Execution Trace Graphing (ETG). In this graph each node represents an action firing and directed arcs represent data (or) control dependency between two various action firings. ETG is a directed and acyclic graph.

By using weighted ETG more optimized buffer size can be gained [8],[9],[10]. In this graph, each firing action is represented with its timing information, hence transform it into weighted graph. It consists of two parts

A. Clock-gating strategy
B. Clock enabling circuit

A. Clock – gating strategy:

The following fig.1 shows the clock gating scheme or strategy. The blocks it contains mainly queue, clock enabling circuit and actor. Actor will be chosen based on the application taken. In this the clock enabling circuit controls the clock of the actor. Whenever the output buffer of the block becomes full then the clock should be turned off inorder to show that the block is in idle state. Switching off the block does not show any effect on the throughput because it is in idle state, there by reduces the power dissipation.

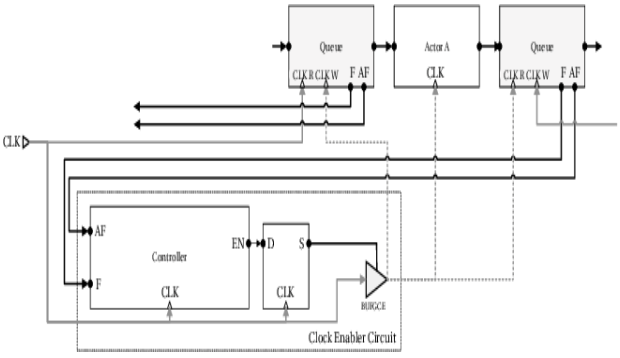


Figure.1 Clock Gating Strategy

Generally RVC-CAL dataflow designs are used for the behavior description, that can be applied to systems that represent execution of process that communicate with asynchronous FIFO buffers. Here the queue blocks should have lossless communication when the actor is clock gated with asynchronous buffers and the design will have different input clock domains.

In the figure shown each queue will have two clocks as inputs, CLKW and CLKR. CLKW is for taking data and CLKR is for producing data. And there are two output ports for the queue. They are Almost Full(AF) and Full (F). From the queue1, the data is given as input to the actor and from the actor the output data is collected by queue2, from this queue2 is readout. When the output of the actor is full, then queue1 will be stopped. To the queue1 the enabling circuit will provide the clock, which will turn off the actor when it is full, hence saving the power. From the fig.1 shown above, the input to the actor is connected to the clock enabling circuit. The clock buffer BUFGCE input clock should be connected to a flip flop to achieve glitch-free clock gating [7].

B. Clock enabling circuit:

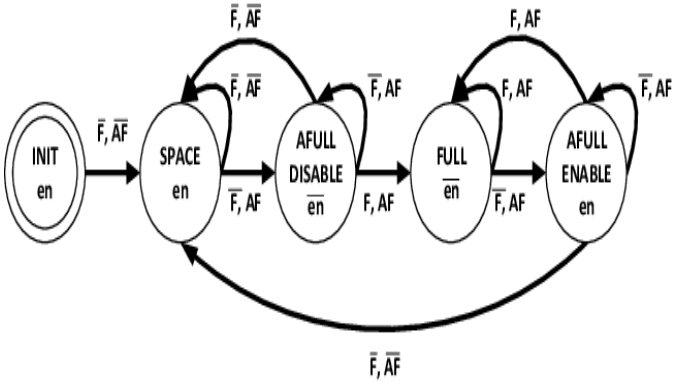


Figure.2 Clock enabling circuit implemented as Finite State Machine

The clock enabling circuit is shown in above fig.2. The shown fig. is implemented as a finite state machine (FSM).

FSM is having a clock, a reset, input Full(F), input Almost Full (AF) and an output enable (EN). The Almost Full AF input becomes active high only when there is one space left to full in its FIFO queue [5]. The above shown FSM has 5 states.

S= {INIT , SPACE , AFULL_DISABLE , FULL , AFULL ENABLE }

The clock-enabling controller circuit starts with INIT state and maintains the ENABLE EN output at active HIGH until F and AF become active LOW. The active high ENABLE EN is maintained during SPACE state. As soon as a queue becomes full, the state will change to AFULL_DISABLE. Here the queue gets disable. In this case, EN output becomes active LOW.

An approach is used in this state as BUFGCE a clock buffer disables the output clock from high-to-low level edge. The enable clock entering BUFGCE must be synchronised to input clock. When a token is taken from the queue, it goes to the AFULL_ENABLE and activates the clock. Then based on the output of buffer Full F (or) Almost Full AF , the state changes to Full (or) to SPACE state.

The user can choose the actor to be clock gated by using mapping configuration. For this, an attribute is given to each actor. The outputs of FIFO queues F and AF are connected to the clock enable controller circuit , if an actor is selected for clock- gating. Output of the queues can be connected directly to a queue (or) through fanout. In the first case, actor output is directly connected to queue without fanout. In the second case, the controller results are connected to other port. In this case one of the fanouts in the queue is full and the fanout should command the actor , but not to produce a token.

III. EXPERIMENTAL RESULTS

This section gives the power reduction gain of the previously mentioned methodology, which is evaluated by applying it to a video decoder design. There are many RVC-CAL applications for dataflow programs [11]. INTRA MPEG4 simple profile decoder is one of the applications. Due to the limitation of number of clock buffers in Xilinx FPGA the design selected was redesigned to result in 32 actors.

MPEG4 video coding is characterized by its scalability and high flexibility. It is a method of compression of audio and the visual digital data. Here the Intra MPEG 4 Simple Profile (SP) description consists of 32 actors. It is a 4:2:0 decoder consists of 8 blocks. Out of these blocks, 4 blocks are luminance (Y) and the other are chrominance U and V, for each two blocks. The parser block contains the bitstream and variable length decoding process, in which U, V and Y are used for texture implements (Tex Y, U,V). The texture decoding consists of variable length coding (VLC), inverse scan test of DCT, inverse quantization, inverse DCT. Whereas, MOT Y, U, V are used to realize the motion compensation stage. Due to the nature of experiments, the MOT stage consists of only one residual error actor. For each queue in the decoder, the minimum queue size is determined [4] by using TURNUS profiler.

For hardware experiment and evaluation, VC707 FPGA kit was used. The code was generated by xronos and was synthesized by Xilinx XST synthesizer. Inorder to produce a netlist, synthesis, routing and placing were applied. The final netlist generated was simulated inorder to extract the Switching Activity Information File(SAIF) of the design. After that, the Xilinx power analyzer was used to get power consumed by using design constraints, design netlist and SAIF inputs also.

The following Table I shows the synthesis results of Intra MPEG4 SP decoder with and without clock-gating.

Logic utilization	Non clock gated	Clock gated	Available
Slices	9214	12776	607200
DSPs	18	18	2850
BRAMs	7	7	1030
LUTs	21499	25126	303600
Max freq.	109	109	-

Table I : Synthesis results with and without clock-gating

A safe option when finalizing your figures is to strip out the fonts before you save the files, creating “outline” type. This converts fonts to artwork what will appear uniformly on any screen.

The above table I shows that clock gated decoder use more slices than non-clock gated. Here the clock gating needs only 15% more than LUTs. A 50 MHz clock is given as synthesis constraint. The Table II shown below gives the power consumption of Clock Gating strategy. In this, two tests are considered. Clock gating enabled and disabled [5], when decoding at maximum output.

Clock gating	Disabled(mW)	Enabled(mW)
Actors clock	58	43
Logic	25	24
Signals	42	41
Clocks	94	80
leakage	242	242
total	403	387

Table II : power consumption when clock gating enabled and disabled

From Table II, the actor clocks show only the power consumption of the actor. Where as, the clock will consist of 50 MHz clock net enabling of clock nets. As a result, the actor clocks consume 26% less power due to clock gating and due to decoder running at high speed, the activation rate of signals and logic result in decrease of 4% in total power.

As from Table I, the maximum decoder output rate is 350 frames/second, for QCIF image of 176x144 pixels. Here the decoder is throttled, such that to decode only 30 images/second for two resolutions. They are CIF(384 x 288 pixels) and QCIF.

The following fig.4 shows (a) the actor power consumption and (b) shows activation rate of actor for each actor clock. The activation rate shows that some of them has an activation rate less than 10%. From this, the power consumption on clocks has reduced by 53.7% for QCIF and 47.6% for CIF resolution. The decoder consumes 54mW less for CIF resolution and 59mW less for QCIF resolution when compared to overall power consumption. From 31 actors, 15 will be always ON, this implies that the 15 actors will never fill up their output. Further, the actors which are not needed for this methodology should find out and eliminate the installation of unwanted additional logic.

In the case of **power saving efficiency and bandwidth,** the decoder will be throttled from 0 to 90% [5] by simulating channel with different consumption rates. It is a case where clock gating is applied for a general application. It is shown in fig.3 below.

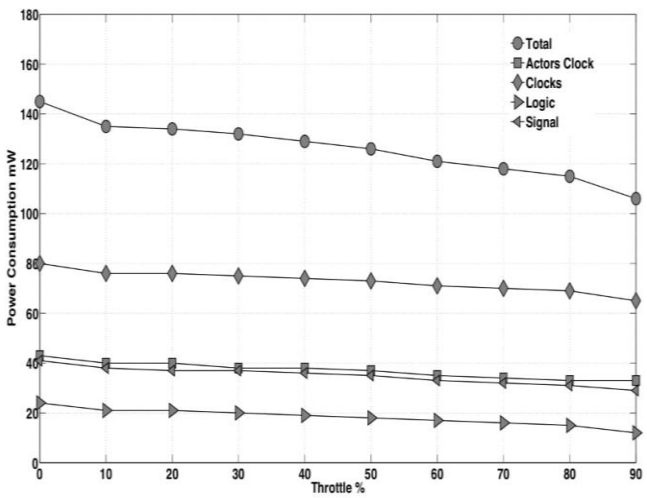
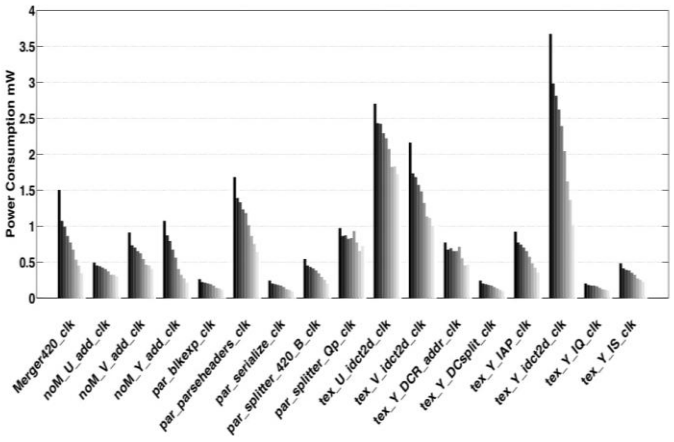


Figure.3 power consumption of clocks, signals, logic and total dynamic power consumption of INTRA MPEG4 SP decoder

From the figure shown above, it is depicted that the total dynamic power has reduced from 145mW to 106mW, a total reduction of 27% power.

The dynamic power of clock gating has been reduced by 34% when compared to non-clock gating. From fig.4 it is shown that, the 15 actors data has been removed due to their activation rate being higher than 99%.



.

Figure. 4(a) Actor clock power consumption

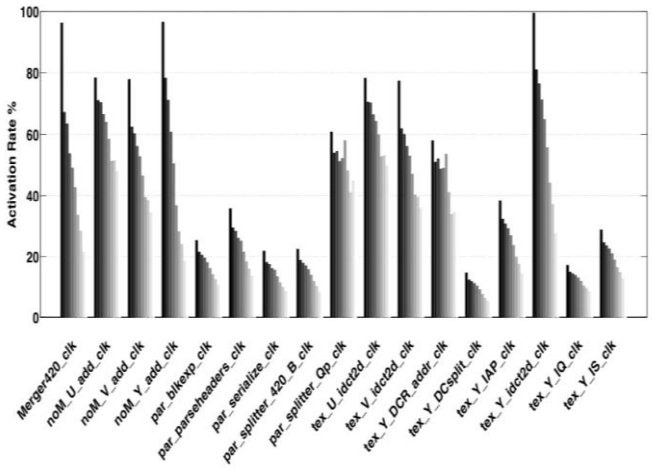


Figure.4(b) Actor clock activation rate

The clock activation rates of actors will be decreased with increase of throttle(except for two cases par_splitter_QP_clk & tex_Y_DCR_addr_clk where power increased slightly). Here, the decoder used is [5] YUV 420. When the power reaches 60%, the chrominance decoding remain active where as, the decoder throttles luminance. It is also occurred during a behavioral simulation in ModelSim.

IV. CONCLUSION

This clock-gating (CG) methodology can be applied to any application inorder to reduce power consumption and reduces the effort during design process at dataflow level. The clock gating(CG) logic is introduced at synthesis stage of HLS design flow. The important component in clock gating is clock enable controller, which is used to give the additional clock. The results show that, when the main block is inactive due to the disabled clock it reduces the switching power. This methodology is very useful where the power dissipation is major challenge.

This method is a simple and effective method inorder to recover power from an idle state. Further developments are necessary inorder to develop the tools used for complex applications onto the limited number of clock domains for more implementations.

V. REFERENCES

[1]. Massoud Pedram, "Power minimization in ic design: Principles and applications," ACM Trans. Des. Autom. Electron. Syst., vol. 1, no. 1, pp. 3–56, Jan. 1996.

[2]. QingWu, M. Pedram, and XunweiWu, "Clock-gating and its application to low power design of sequential circuits," Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on, vol. 47, no. 3, p. 415–420, Mar 2000.

[3]. G.E. Tellez, A. Farrahi, and M. Sarrafzadeh, "Activity-driven clock design for low power circuits," in Computer-Aided Design, 1995. ICCAD-95. Digest of Technical Papers., 1995 IEEE/ACM International Conference on, Nov 1995, pp. 62–65.

[4]. M. Canale, S. Casale-Brunet, E. Bezati, M. Mattavelli, and J. Janneck, "Dataflow programs analysis and optimization using model predictive control techniques," Journal of Signal Processing Systems, pp. 1–11, 2015.

[5]. Endri Bezati, Simone Casale-Brunet, Member, IEEE Marco Mattavelli, and Jorn W. Janneck, Member, IEEE "clock-gating of streaming aplications for energy efficient implementations on FPGAs" DEC-2015.

[6]. H Mahmoodi, V Tirumala shetty, M Cooke and K Roy," Ultra Low-Power Clocking Scheme Using Energy Recovery and Clock Gating," *in IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 17, no. 1, pp. 33-44, Jan. 2009.

[7]. Xilinx, Analysis of Power Savings from Intelligent Clock Gating, August 2012, XAPP790.

[8]. S. C. Brunet, M. Mattavelli and J. W. Janneck, "Buffer optimization based on critical path analysis of a dataflow program design," *2013 IEEE International Symposium on Circuits and Systems (ISCAS2013)*, Beijing, pp. 1384-1387,2013.

[9]. B. Ghavami and H. Pedram, "High performance asynchronous design flow using a novel state performance analysis method, *Computational Electrical Engineering*, vol.35, no.6, pp. 929-941, November 2009.

[10]. S Suhaib, D Mathaikutty, and S Shukla, "Dataow architectures for GALS," *Electron. Notes Theor. Comput. Sci.*, vol. 200, no. 1, pp. 33–50, 2008.

[11]. "Open RVC-CAL Applications," 2014, http://github.com/orcc/orc-apps, accessed 25-February-2014.

TRANSITION INVERSION BASED LOW POWER DATA CODING SCHEME FOR BUFFERED DATA TRANSFER

Suhasini Parunandi
PG Student
psuhasini11@gmail.com

Malla Reddy College of Engineering (MRCE) Dulapally ,Telangana ,India

Abstract

In this work the authors propose a data coding protocol that leads to power reduction for block data transfer in off-chip buses. I/O pads driving off-chip buses contribute to a major portion of power dissipation in chips. Also, block data transfer is preferred in most systems like caches, DMA etc. In this proposed work, the prior knowledge of the block of data to be transmitted, when it is stored in the buffer, is exploited in a serial fashion to reduce transitions on every bus line. Statistical analysis shows up to 31.9% reduction in transitions. Benchmark results show that it leads to 29% reduction in power consumption. The technique provides added error detection on the lines of parity bit technique, with similar average error detection capability.

1. Introduction

Increased integration and higher operating frequencies compound the problem of power dissipation in VLSI chips. One of the main contributors to power consumption is switching activity on the high-capacitance lines of an interconnection system, especially off-chip buses. Buses constitute an important resource for addressing and data transfer in the implementation of most electronic systems. The fact that the power consumed at the I/O pads accounts for a significant fraction of the total power consumed in VLSI systems has been independently established by many researchers [1-3].

Error detection is also of equal importance, as buses are more prone to error due to temperature variations, interference from neighboring sources, and ageing etc. Thus a low power data coding scheme supporting error detection is desired. Existing techniques make use of the data as it is placed on to the bus. This leads to a delay which can severely affect the operating frequency of the circuit. The proposed encoding technique, which is based on serial transition inversion, works on blocks of data for power reduction in off-chip buses. Numerous bus transmission protocols deal with blocks of data rather than individual data words. This is evident in DMA transfers and cache lines which are widely used in computer systems. In block data transfer the latency in data coding can be hidden since the block will be transmitted only after it is filled up. This brings the pipeline approach to the coding technique. Since Off-chip buses consume more power, keeping the encoding circuitry before the I/O pad reduces power consumed by it. The proposed technique is also called transition inversion in the following sections

2. Related work

One of the most often cited encoding methods is the bus-invert method [1]. Bus-invert selects between the original and the inverted pattern in a way that minimizes the switching activity on the bus. The resulting patterns together with an extra bus line (to notify whether the address or its complement has been sent) are signaled over the bus. Musoll et al. proposed the working zone method [3]. Their method takes advantage of the fact that data accesses tend to remain in a small set of working zones. Other encoding techniques include Asymptotic Zero-Transition Encoding [2], Gray coding (mostly for addresses), and other application specific encoding techniques [9,10]. Most of the existent techniques make use of the repeating patterns in address buses to reduce address bus transitions [6,8,9]. Frequent Value encoding (FVE) is another technique proposed in [13] also results in a significant reduction in transitions, but has not been considered here, due to the overhead involved.

There is no existent literature on bus coding methodologies for block data transfer, other than Serial Bus Invert [14], which encodes blocks of data rather than individual data words. Also a transition inversion

based encoding technique [12] was developed for serial buses that serve as a basis for the proposed technique. The technique proposed in this work is compared with Serial Bus Invert, serial gray coding [15], and transition signaling [11], which is the reverse of gray coding. The last technique mentioned is similar to the algorithm proposed by us, but differs in the decision making process.

3. Proposed technique

In block data transfer, data is generally loaded onto a buffer and then is put on the bus. Each line in the bus is a serial line that will transmit one particular bit position of all data words that are put on the bus. Before transmission, the number of transitions on a line is counted. This can be done by a simple XOR gate between consecutive bits and counting the number of '1's. If the number of transitions is more than half the number of data words, the transitions states between the bits can be inverted. Each transition is made as a non-transition and vice versa. If not, the bit stream is transmitted as such. In case transition inversion is needed, the scheme operates by observing the transition states between any 2 bits and setting the encoded second bit to be the same as the previous encoded bit if there is a transition. If there is no transition, the previous encoded bit is inverted. The decision bit signifying transition inversion is transmitted before transmitting the encoded data. This has to be done on all lines. Since the data is sent as a block, the extra bit on each line will signal for all the data words.

The transitions in the bit stream, transmitted on a line, can be reduced by the aforementioned scheme. Each line is processed independent of each other. If there is a need for transition inversion, then the following steps are followed to obtain encoded data. Let the data bit that is to be transmitted next be b_d and previous data bit be b_{dp}. The previous transmitted bit is b_{tp}. The next transmitted bit will be

bt = btp if bd ≠ bdp
= !btp if bd = bdp

In receiver the reverse logic needs to be applied. When the bit stream has been signaled as modified, then the following steps are followed to decode data. The previous and current received bits are assumed to be 'brp' and 'br' respectively. The previous decoded bit is assumed to be bdp. The current received bit will be

bd = bdp if brp ≠ br
= !bdp if brp = br

The encoding and decoding is done on the fly, to reduce performance losses. For example in the bit stream 10101011, the number of transitions is 6. Thus this stream is to be modified according to the algorithm described above. The first bit is transmitted as such, without any change. This is described in Table 1. The encoded data has only one transition.

Table 1: Sample coding process

Key: NT – No Transition, T – Transition, NC-No Change

Bit No.	1	2	3	4	5	6	7	8
Bit stream	1	0	1	0	1	0	1	1
Transition State	NC	T	T	T	T	T	T	NT
Encode state	NC	NT	NT	NT	NT	NT	NT	T
Encoded Bit stream	1	1	1	1	1	1	1	0
Decode state	NC	T	T	T	T	T	T	NT
Decoded bit stream	1	0	1	0	1	0	1	1

A. Statistical Analysis of the Algorithm

A statistical analysis of the serial transition inversion (STI) algorithm has been carried out in two independent methods taking buffer depths of 8, 16, 32, and 64. The first analysis considers all combinations of the word and determines the original and modified transitions in the datasets. It is essentially a brute force approach. The data considered was a uniform distribution of all possible data patterns that is likely to be transmitted over the bus. For example, considering a buffer depth of 8, the number of possible data patterns of one bit stream is 256. The number of transitions in these data patterns was calculated along with the number of transitions in the data pattern after being modified, using the proposed algorithm. The second analysis was an analytical one. For an N-bit system, where the transitions are taken between the consecutive bits, maximum of (N-1) transitions are possible in the bit stream. Number of possibilities of 'i' transitions= $^{(N-1)}C_i$. Let T_{org} and T_{mod} be the number of transitions in the original data patterns and the number of transitions in the modified data patterns respectively. These entities can be calculated as follows: Total number of transitions T_{org}

$$\sum_{i=0}^{N-1}\left((N-1)_{C_i} * i\right) \quad (1)$$

Transition inversion is done when the number of transition is more than or equal to N/2. The number of transitions in the modified data will be (N-1-i) for 'i' transitions in the original data.

$$T_{mod} = \sum_{i=0}^{\frac{N}{2}-1}\left((N-1)_{C_i} * 2 * i\right) + \sum_{i=\frac{N}{2}}^{N-1}\left((N-1)_{C_i} * 2 * (N-1-i)\right) \quad (2)$$

The average reduction can be calculated by taking the difference between the number of transitions in the modified data patterns, given by equation (2) and unmodified data patterns, given by equation (1). The reduction figures would be smaller if the decision bit is also considered. The results obtained by both the methods agree with each other and are shown in Table 2.

Table 2: Statistical Percentage Reduction

Word Length	8 Bits	16 Bits	32 bits	64 bits
% Reduction in transitions	31.25	20.95	13.54	9.78

B. Power Analysis of the Algorithm

The overall power reduction consists of the power reduction achieved by the transition reduction minus the power consumed by the extra circuitry. Unmodified dynamic power consumed by I/O pads is given by

$$p_{org} = \frac{1}{2} V_{dd}{}^2 C_T f \alpha \qquad (3)$$

Where, Vdd, f, Ct, α represent drain voltage, frequency of operation, line capacitance, switching activity respectively. If the power dissipation of the extra circuitry required for the coding process is taken into account then the equation given above has two extra terms on the right hand side, the encoder and decoder power dissipation respectively.

C. Performance Analysis of the Algorithm

The transition inversion algorithm needs an extra bit to be transmitted before the start of the block of data on all lines. This leads to a decrease in bandwidth utilization. For a system with buffer depth of 8, 9 bits are transmitted on one line. Therefore a frequency increase of 9/8 will be needed to maintain the same bandwidth utilization. The corresponding power consumed by I/O pads will increase linearly.

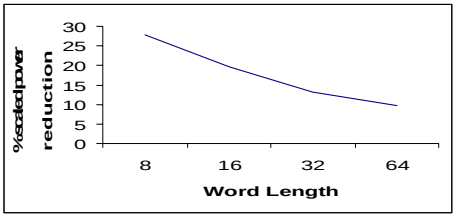


Figure 1: Effects of frequency scaling with word length

A performance metric is defined to take into account the scaling of the frequency and the reduction in transitions, and is calculated as their product. The variation of this parameter with word length is shown in Figure 1. Bus invert leads to a reduction in bandwidth since it poses a delay in putting the encoded data. By following a depth based approach where most delays are hidden, bandwidth need not be reduced.

D. Error Detection Analysis of the Algorithm

The proposed algorithm's propensity towards reducing the number of transitions can be used for detecting errors. This can be done by determining if the number of transitions in the received bitstream is more than half the bitstream length. If this count is more than half the bitstream length, the incoming data is incorrect. The proposed technique is compared with parity bit technique, as both have similar overhead i.e. addition of one bit to the bitstream. The parity bit detects all odd bit errors, but misses even bit flips, whereas, transition inversion can detect a certain percentage of any number of bit errors. Error analysis has been done by considering all combinations of the given word length that are transmitted over the bus. For transition inversion coding, the possible combinations of the word are those in which the number of transitions is less than 4. All the combinations of bit errors right from one bit error to 8 bit errors have been checked for both the proposed technique and parity bit technique. The result of this analysis is shown in Table 3.

Table 3: Error Detection Analysis

No. of Bit errors	% of errors detected	
	Parity Coding	Proposed Technique
1	100	31.25
2	0	44.64
3	100	52.68
4	0	55.71
5	100	52.68
6	0	44.64
7	100	31.25
8	0	0

If all the bits are in error, then neither technique can detect the error, as in the proposed technique if all bits are flipped, the number of transitions remains the same. Calculation of statistical averages over the entire range of bit errors shows that the proposed technique and parity bit technique both have the same value of 50.2%. The average is calculated as the ratio of total number of errors detected to the total number of errors possible on the line. Thus the proposed technique can be used as a hint to upper layers of communication that an error has occurred since it cannot reliably detect all errors.

4. Implementation strategies

In most block systems, the data buffer is present just before the transmission part. The core logic puts the data inside the buffer from one side and the transmission happens from the other side.

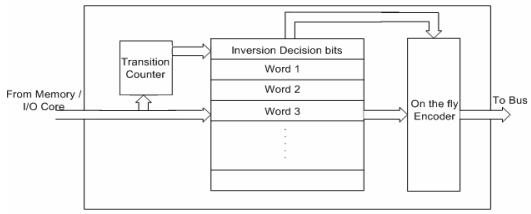


Figure 2: High Level Architecture

The transition counting happens when the data is being filled up in the buffer. The transition inversion decision is made depending on the count of the transitions and encoding done based on it. The bit stream is encoded on the fly as the data is put on the bus, as shown in Figure 2. In the receiver the decoder has to decode the incoming bit stream and recover the original data.

A. Decision Circuit and Encoder

The decision circuit is a simple XOR gate between consecutive bits of the input bit stream as shown in Figure 3. The counter needs to count only up to half the number of maximum transitions.

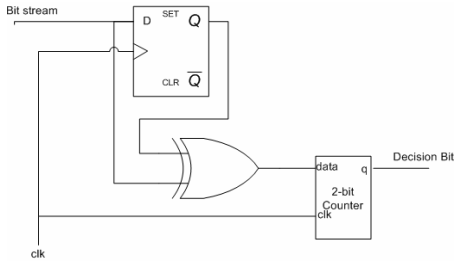


Figure 3: Decision Circuit (Transition Counter)

This circuit can also be implemented with double edge triggered circuits to further optimize at the encoder stage.

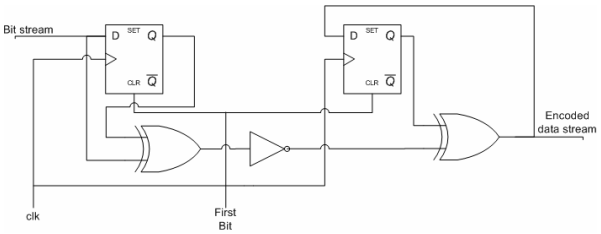


Figure 4: Encoder Circuit

The transition counter works in parallel to buffer loading, and is thus masked. The on the fly encoder is shown in Fig 4.

B. Decoder

The decoder shown in Fig 5 performs XOR between consecutive bits to determine transition state, inverts the received bit if required to recover the data.

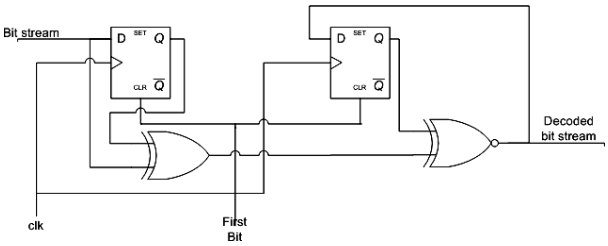


Figure 5: Decoder Circuit.

C. Complexity Analysis

The main components of both systems are the decision circuit, encoder and decoder.

Decision circuit: With increase in bus width, the space complexity increases exponentially ($O(N^2)$) for bus invert decision circuit. A comparable parameter in the proposed technique is buffer depth which leads to a linear increase (O(N)) in circuit complexity. Time complexity increases linearly (O(N)) in bus invert, while in the proposed technique it is constant(O(1)).

Encoder and Decoder: With increase in bus width, the bus invert encoder/decoder circuit complexity increases linearly (O(N)). For the proposed technique, the circuit complexity is constant (O(1)). Time complexity is constant (O(1)) in both bus invert and the proposed technique.

5. Experimental results

For experimental analysis, the algorithm was applied on random image data and SPEC2000 benchmark binaries.

Random Image Data:

For this analysis of the algorithm seven images were taken and their RGB values were ran through the algorithm. The images were a mix of both smooth and detailed features. The results are tabulated in Table 4. These do not include the power dissipated by the encoder and decoder circuitry.

Table 4: A comparison of transition reduction for Bus invert and the Proposed Technique

#	Original no. of transitions	Bus Invert Coding		Proposed technique	
		transitions	% reduction	transitions	% reduction
1	120160	86296	28.18	72212	39.9
2	127770	94776	25.82	85454	33.11
3	74666	61746	17.3	53502	28.34
4	165678	119578	27.83	119908	27.62
5	111909	81645	27.04	70978	36.58
6	66189	49251	25.59	46769	29.34
7	159620	121466	23.9	114163	28.48

It is clear from the above table that STI performs much better than bus invert.

With SPEC2000 benchmarks:

SPEC2000 benchmark binaries traces were run with the proposed technique and compared with bus invert and gray coding. The 26 binaries were run with varying the buffer depth and bus widths with the values 8,16,32,64. The averages for a given combination of bus width and buffer depth were taken and have been plotted. The results for the proposed technique and bit invert are showed in Figure 6 and 7 respectively.

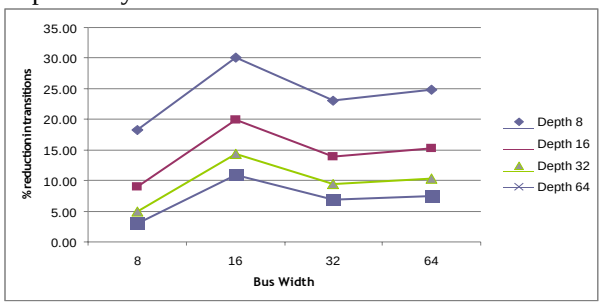


Figure 6: Transition reduction in Transition Inversion

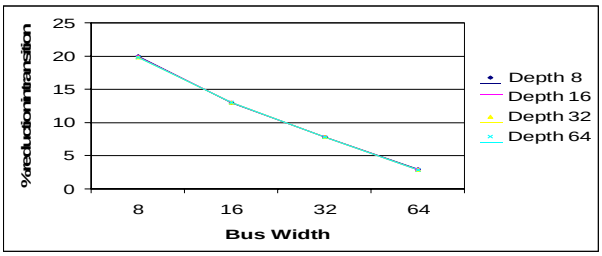


Figure 7: Transition reduction in Bus invert

It can be observed that the buffer depth does not make any changes to bus invert. Also with increase in buffer depth, the transition reduction reduces for the proposed technique. A similar observation can be made for bus invert when bus width is increased.

With increasing bus widths in present VLSI systems, the proposed technique will perform better. The increase in buffer depth leads to a lesser reduction. It can be offset by splitting the block into sub-blocks of smaller depths. Depending on the system, a compromise between power reduction and bandwidth utilization can be found. The benchmark files were also run using Gray Coding technique which is the reverse of transition signaling. The results are shown in Figure 8.

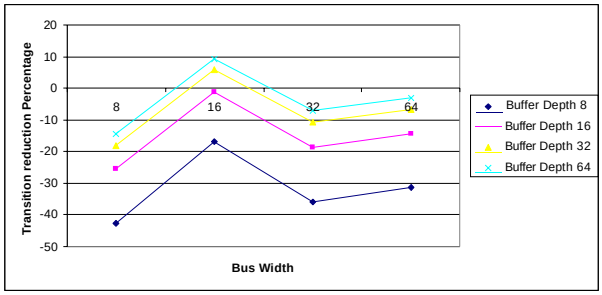


Figure 8: Transition reduction in Gray Coding

It can be seen that there is not much reduction in transitions, with some of the data points showing an increase in the number of transitions. Gray code does not show any reduction in transition since it is an N-bit to N-bit mapping. Whatever data tuples are in input set are exactly the same in the output set. So it is not possible to get any benefits out of Gray code. The same is the case with transition signaling.

B. Overall Power Analysis

The proposed system was designed in RTL and analyzed with Synopsys synthesis tools. A bus operating at 100MHz was assumed with its I/O voltage levels at 3.3v. The internal circuitry was modeled on 180nm process technology. The circuitry was simulated by feeding the SPEC2000 benchmark trace files as input for a buffer depth of 8. The power consumed by the circuitry is shown in Table 5.

Table 5: Power dissipation of encoding/decoding circuitry for transition inversion

Circuitry	Decision circuit	Encoder	Decoder
Power consumed	28.7µW	28.9µW	28.6µW

Assuming the parameters stated above and the activity factor for the benchmarks to be 0.5, the power was found to be 27.23mW. The reduction in power consumption is linearly dependent on the activity factor reduction. A reduction of 30% activity leads to a reduction of power by 8.17mW. The total power consumed by the extra circuitry is 86.2 µW leading to a

net power reduction of 8.08mW which corresponds to 29.7 % reduction in power.

C. Overall Delay Analysis

The proposed technique does not involve circuitry of multiple stages thus leading to less delay. The delay performance of the proposed technique and bus invert in terms of propagation speed is compared in Table 6.

Table 6: Comparison of encoder delay in Bus Invert and Transition Inversion

Technique	Proposed Technique	Bus Invert
Delay	1.2ns	3.3ns

For calculating the delay due to the proposed technique only the encoder is considered. The decision circuit is not taken as it will be part of the buffer loading delay and will not contribute to encoding. The decision circuit delay was found to be 0.2ns since it involves only the XOR gate. The flip flops delays will be masked by the sequential loading thus giving a pipelined approach. For the bus invert technique, the decision circuit delay is also taken into account because encoding has to be complete before the next data word arrives. Thus before the next data arrives the counting of the hamming distance should have been done and the data encoded. Overall the encoding delay of the proposed technique is considerably lesser compared to that of the bus invert.

6. Conclusions

In this paper an encoding technique has been presented that reduces power dissipated on off-chip data buses for block data transfer. The technique involves inverting the transition states on every line of the bus if the transitions exceed the number of non-transitions. The modification status is signaled as an extra word, thus avoiding the use of an extra line. The average reduction obtained in terms of transitions is 31.9% while the net power reduction after the extra power circuitry is taken into account is 29.7%. This is achieved without using an extra bus line. The compromise is in bandwidth utilization which can be adjusted by choosing a proper block length. This technique can also be applied to synchronous serial buses. The presence of a parity bit like error detection mechanism in addition to low power gives it an additional advantage.

7 References

[1] M. R. Stan, W. P. Burleson. “Bus-Invert Coding for Low Power I/O”, IEEE Transactions on Very Large Integration Systems, Vol. 3, No. 1, pp. 49-58, March 1995.

[2] L. Benini, G. De Micheli, E. Macii, D. Sciuto, C. Silvano. “Asymptotic Zero-Transition Activity Encoding for Address Buses in Low-Power Microprocessor-Based Systems”, IEEE 7th Great Lakes Symposium on VLSI, Urbana, IL, pp. 77-82, Mar. 1997.

[3] E. Musoll, T. Lang, and J. Cortadella. “Working-Zone Encoding f or reducing the energy in microprocessor address buses”. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 6, no. 4, Dec 1998

[4] E. Musoll, T. Lang, and J. Cortadella. “Exploiting the locality of memory references to reduce the address bus energy”, Proceedings of ISPLED, pp. 202-207, Monterey CA, August 1997.

[5] M. Stan and W. Burleson. “Low-power encodings for global communications in CMOS VLSI”, IEEE Trans. on VLSI Systems., pages 444–455, 1997.

[6] L. Benini, G. D. Micheli, E. Macii, M. Poncino, and S. Quer. “Power optimization of core-based systems by address bus encoding”, IEEE Transactions on Very Large Scale Integration, 6(4), Dec. 1998.

[7] Tomas Akenine-Moller, Jacob Strom. “Graphics Processing Units for Handhelds”, Proceedings of the IEEE Vol. 96, No. 5, pp. 779-789 May 2008.

[8] L. Benini, G. D. Micheli, E. Macii, D. Sciuto, and C. Silvano. “Address bus encoding techniques for system-level power optimization”, In IEEE Design Automation and Test Conference in Europe, pp 861–866, Paris, Feb. 1998.

[9] Giuseppe Ascia, Vincenzo Catania, Maurizio Palesi. “A Genetic Bus Encoding Technique for Power Optimization of Embedded Systems”, J.J. Chico and E. Macii (Eds.) PATMOS 2003, LNCS 2799, pp. 21–30, 2003. Springer-Verlag Berlin Heidelberg 2003.

[10] C. Su, C. Tsui, and A. Despain. “Saving power in the control path of embedded processors”, IEEE Design and Test of computers, 11(4):24–30, 1994

[11] Wei-Chung Cheng, Massoud Pedram. “Memory Bus Encoding for Low Power: A Tutorial”. ISQED 2001: 199-204

[12] Abinesh R., Bharghava R., M.B. Srinivas, “Transition Inversion Based Low Power Data Coding Scheme for Synchronous Serial Communication”, isvlsi, pp.103-108, 2009 IEEE Computer Society Annual Symposium on VLSI, 2009

[13] Jun Yang, Rajiv Gupta, Chuanjun Zhang. “Frequent value encoding for low power data buses”. ACM Trans. Design Autom. Electr. Syst. 9(3): 354-384 (2004)

[14] Saneei M, Afzali-Kusha A, Navabi Z. “Serial Bus Encoding For Low Power Applications”, International Symposium on System-On-Chip, pp. 1-4, November 2006.

[15] Kangmin Lee, Se-Joong Lee, Hoi-Jun Yoo. “SILENT: serialized low energy transmission coding for on-chip interconnection networks”. ICCAD 2004: 448-451

AN IOT BASED SOLAR VEHICLE WITH OBSTACLE AVOIDANCE NIGHT SPEED LIMITTER AND ACCIDENT DETECTION USING GSM AND GPS

MAMIDI PRUDHVI RAJ[1]

[1]*PG Student, Dept. of ECE*, MRCE, Hyderabad, India, Email: prudhvi7269@gmail.com

ABSTRACT:

The purpose of our work is to discover the vehicle accident location through sending a message using a system which is placed within automobile system. Moreover there is a rapid rise in the event of the Roadway accident. This paper is about a system which is created to automatically detect a mishap as well as inform the nearest health centers and medical solutions about it. This system can also locate the area of the mishap so that the clinical solutions can be directed right away in the direction of it. The objective of this paper is to build up a Car unintended tracking system utilizing RESONANCE SESNOR, GPS as well as GSM Innovation. The system can be adjoined with the car alarm system and also notify the owner on his mobile phone. This discovery and also messaging system is made up of an Aurdino receiver, Microcontroller and a GSM Modem. GPS Receiver obtains the area information from satellites in the form of latitude and longitude. The GSM modem sends out an SMS to the predefined mobile number and informs about this accident. This enable it to keep an eye on the mishap circumstances and it can immediately signals the police/ambulance service with the location of mishap. The entire modules are can run with help of solar energy with battery power storage system (BESS).

Keywords: GSM, GPS, Vibration sensor, accident, latitude and longitude, location.

1. INTRODUCTION:

The usage of vehicle mobiles has boosted linearly over the past years, which boosted in the danger of human life. This is because as a result of the inadequate emergency facilities. In this paper we are using an alarm system which assists in improving the emergency situation system of the crash system. This system identifies the crash event and the collaborated of the crash are messaged to the rescue team. A changing system is used turn off in case there is no origin. The Crash is discovered with the help of RESONANCE SESNOR Sensing Unit and Vibration Sensing Unit. The Angle in which the automobile has rolled off is suggested through a message. This Application helps in supplying possible service to the bad emergency situation assists in. The function of the project is to locate the car where it is and also locate the vehicle by means of sending a message using a system which is placed inside of automobile system Most of the moments we may not have the ability to find crash area because we don't recognize where crash will happen. In order to give treatment for damaged individuals, initially we require recognizing where the crash occurred through place tracking and also sending a message to your associated one or to the emergency situation solutions. So in this job we are making use of the standard microcontroller for inexpensive as well as additionally for easy understanding. Here we made use of setting up programs for far better accuracy and GPS as well as GSM components which assists to map the car anywhere on the world. The specific area of the automobile is sent out to our remote devices (cellphones) making use of GSM modem. We are in the procedure of solving this issue by suggesting a reliable service and to lower the loss of lives as long as possible. In our concept, the layout of the system

assist us to find crashes in significantly less time and also move the basic information to the emergency treatment centre within a few secs covering the geographical works with, the moment and the angle where the vehicle had met with a mishap. This sharp message is sent to the rescue team (rescue) and the family within the brief duration. This real time application saves numerous useful lives. The message is sent with the GSM component and the area of the basic idea is to center the vehicle system by obtaining the live placement of the vehicle through GPS as well as send out the details via GSM component by means of SMS service with an added attribute of GPRS transmission to the tracking center via usage of web [M.AL-Rousan, A. R. AI-Ali as well as K. Darwish et al, 2004] Using microcontroller, this project has actually been designed. It made use of EEPROM to keep the phone numbers. Crash is discovered with the help of the GPS component. The accident can be identified precisely with the help of both Micro electro mechanical system (VIBRATION SESNOR) sensing unit and also vibration sensor. The Angle of the surrender of the cars and truck can likewise be understood by the message via the RESONANCE SESNOR sensing unit.

2. RELATED STUDY:

Currently requirements, we cannot find where the mishap has happened as well as for this reason no details pertaining to it, causing the death of an individual. The research work is taking place for tracking of the vehicle also in dark clumsy locations where there is no network for getting the signals. In literary works, a number of strategies to provide security as well as security via monitoring the lorry's real time precise positioning and also info utilizing different technology-nologies have been proposed. An excellent survey of using GPS, GSM

and GIS has actually been given in [IoanLita, Ion BogdanCioc, Daniel AlexandruVisan et alia, 2006] and also [Mrs. Ramya Kulandaivel, P.Ponmalar, B.Geetha, G.Saranya et alia, 2012] The general mechanism is to give the live geographical placement of an automobile utilizing GPS receiver and send this info to GSM facility via configurable software, this is all done by the monitoring facility which is functioning as a control unit that is attached not only by an optical wire but likewise linked wirelessly with TCP/IP protocols. The monitoring center disperses the data to the client in a reasonable format as well as it also keeps the travelling documents as well as presents the actual time information concerning automobile on digital map with GIS system [IoanLita, Ion BogdanCioc, Daniel AlexandruVisan et al, 2006] Another strategy is that automobile terminal consists of a GPS receiver which draws out details about setting with GPS satellites and also sends it via GSM network as well as to the nerve center which reads.

3. WORKING METHODOLOGY:

We propose an intelligent vehicle system for mishap avoidance and making the world a much better as well as refuge to live. Easy IR sensor is reliable for finding human or animals as well as this strategy certainly can conserve lots of lives. Pre collision detection system have to be equipped with mix of different sensors. Discovering humans or pets including challenges will certainly give us a far better service to reduce the death of human in road crash. INFLUENCE OF THE RECOMMENDED SERVICE: Presently criteria, we can not spot where the mishap has actually taken place as well as therefore no info pertaining to it, leading to the fatality of a person. The research study work is taking place for tracking the setting of the automobile even in dark

clumsy areas where there is no network for obtaining the signals. In this job GPS is used for tracking the placement of the vehicle, GSM is utilized for sending the message. Thus with this job application we can find the position of the vehicle where the accident has happened to make sure that we can provide the first aid as early as possible. This task provides lorry crash discovery and sharp system with SMS to the individual defined mobile numbers.

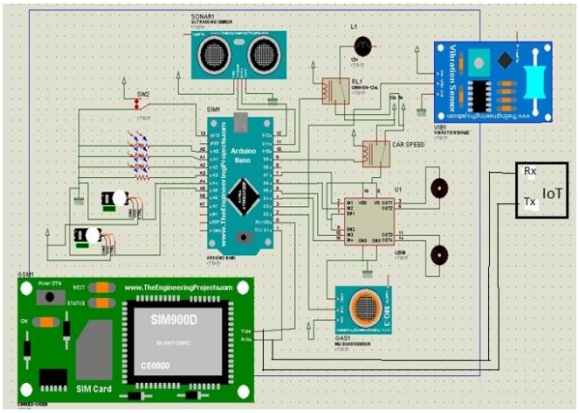


Fig.3.1. Schematic diagram of our proposed system.

When the system is turned on, LED will certainly get on suggesting that power is provided to the circuit. When the IR sensors that we are making use of in our task sense any kind of barrier, they send out disrupt to microcontroller. The GPS gets the area of the vehicle that met a crash and offers the information back. These details will be sent out to a mobile number through a message. This message will certainly be obtained utilizing GSM modem existing in the circuit. The message will certainly give the information of longitude and also latitude worth. Making use of these worth the placement of the vehicle can be estimated.

GPS LOCATION:

The Global Positioning System is a room based worldwide navigation satellite system that supplies reliable location (crash area) and also times anywhere on the world. The GPS satellites act as a referral factor from which receivers on the ground identify their setting. The essential navigating concept is based on the dimension of pseudo ranges in

between the user as well as 4 satellites.

Fig.3.2. GPS module.

ULTRASONIC SENSOR:

The Ultrasonic Sensing unit sends a high-frequency audio pulse and after that times how much time it considers the echo of the sound to mirror back. The sensor has 2 openings on its front. One opening up transmits ultrasonic waves, (like a little audio speaker), the other receives them, (like a little microphone).

Fig.3.3. Ultrasonic sensor.

ALCOHOL SENSOR:

The MQ-3 alcohol gas sensing unit includes overall 6-pins including A, H, B and the various other three pins are A, H, B out of the overall 6-pins we utilize just 4 pins. The two pins A, H are made use of for the home heating purpose and also the various other two pins are made use of for the ground and power. There is a heating unit inside the sensor, which is made up of aluminium oxide, tin dioxide. It has warmth coils to produce heat, and also hence it is used as a heat sensor. The below diagram shows the pin representation and the arrangement of the MQ-3 alcohol sensor.

Fig.3.4. Alcohol sensor.

SOLAR TRACKER:

Trackers straight photovoltaic panels or components toward the sun. These tools alter their orientation throughout the day to adhere to the sun's path to make the most of energy capture. In photovoltaic or pv systems, trackers aid lessen the angle of incidence (the angle that a beam makes with a line vertical to the surface) between the inbound light and the panel, which raises the amount of power the installation creates. Focused solar photovoltaic's and focused solar thermal have optics that straight accepts sunshine, so solar trackers need to be angled appropriately to accumulate energy.

Fig.3.4.Solar dual direction operation.

The Arduino Nano controller is used to control all the sensor modules. A dual axis tracking prototype is developed to capture the maximum sun rays by tracking the movement of the sun in four different directions. The car will move only when the person wears the seat belt and doesn't sip ant alcohol. For this, we have a seat belt and alcohol sensors. The ultrasonic sensor is used to measure the distance if any obstacle comes close to the vehicle and car will turn either left or right automatically. At the time if the vehicle goes with over speed then automatically the headlights of the vehicle goes dim.

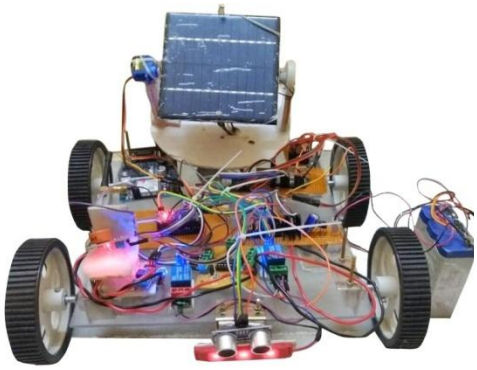


Fig.3.5. Prototype model.

If the vehicle goes in normal speed then headlights of the vehicle are blown brighter, this operation can be done by Relays and switches. If the vehicle met with any accident vibration sensor will sense it, an alert SMS will be sent to people like friends, ambulance, Police through GSM module. By using ESP-12 Wi-Fi module, and GPS module the whole information about the vehicle means its speed, condition all are updated in the Android Application. Motor drivers are used to driving the motors with the same voltage.

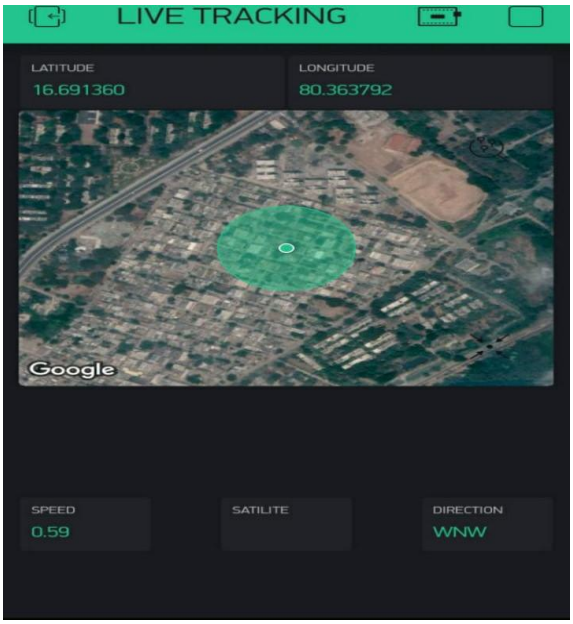


Fig.3.6. Final GPS output.

The listed below number shows that at the point when accident struck the car Vibration Sensing unit, which finds the accident and also subsequently sends out the signals to Arduino. At this point the Arduino takes control and also begins accumulating the collaborates gotten from the controller which are later sent to the Central Emergency Situation Monitoring Terminal by utilizing the GSM Component. After that the alert message will be sent on signed up mobile number through gsm.

5. CONCLUSION:

The proposed system is established to provide the information about the accident occur with location of the crash, prevents the drunk & drive accident. I conclude that my theory was correct, that the angle of the photovoltaic panel dealing with directly at the sunlight made the solar energy automobile go the fastest. The angle of the solar panel will be readjusted with respect to the activity of the sun.The car doesn't relocate up until the chauffeur wear the seat-belt and also doesn't consume any alcohol. If the chauffeur goes fast during night time the head lights of the car will immediately lower as well as get into typical circumstance if the he enters normal speed. If any other automobile comes closer to our automobile instantly an alert message will certainly be given to the driver and progressively the speed of the lorry will certainly be lowered.

REFERENCES:

[1] Vikram Singh Kushwaha, Deepa Yadav, Abuyeed Topinkatti, Amrita Kumari. "Auto Mishap Discovery System utilizing GPS And GSM", Volume 2, Concern 1(Jan-Feb 2015), PP12-17

[2] Nimisha Chaturvedi, Pallika Srivastava. "Automatic Lorry Accident Detection and Messaging System Making Use Of GSM as well as GPS Modem ", Volume: 05 Issue: 03|Mar-2018

[3] C.Prabha, R.Sunitha, R.Anitha. "Automatic Car Mishap Discovery and also Messaging System Utilizing GSM and GPS Modem", Vol. 3, Concern 7, July 2014

[4] Hoang Dat Pham, MichealDrieberg, Chi Cuong Nguyen, "Advancement of car tracking system making use of GENERAL PRACTITIONER and GSM modem ", Seminar: 2013 IEEE Conference on Open Equipment (ICOS).

[5] Lih-Jen Kau, Member, IEEE, and also Chih-Sheng Chen, "A Smart Phone-Based Pockert Loss Crash Discovery, Positioning As Well As Rescue System", Dec 2013.

[6] G. Acampora, D. J. Cook, P. Rashidi, A. V. Vasilakos, "A Study on Ambient Intelligence in Health Care", Proceedings of the IEEE, pp. 2470-2494, Vol. 101, No. 12, Dec. 2013.

[7] Malik Tubaishat, Qi Qi, Yi Shang, Hongchi Shi "Wireless Sensor-Based Traffic Signal Control" IEEE CCNC 2008 process 1-4244-1457-1/ 08.

[8] "Automatic crash notice system using gsm and also gps modems with 3g technology for video monitoring" International Journal of Emerging Fads in Electric and also Electronics (IJETEE) Vol. 1, Issue. 2, March-2013.

[9] E.Krishna Priya, P.Manju, V.Mythra, "IoT Based Lorry Tracking and Accident Discovery System" International journal of Innovative Study in Computer as well as Communication Engineering, (An ISO 3297:2007 Licensed association) Vol.5, Concern 3.

[10] Sri Krishna Chaitanya Varma, Poornesh, Tarun Varma, Harsha; Automatic Car Accident Detection And Also Messaging System Utilizing GPS and also GSM Modems; International Journal of Scientific & Design Research study, Volume 4, Problem 8, August-2013.

DLAU: A Scalable Deep Learning Accelerator Unit on FPGA

VAJJA PARMESH,PG student,Dept of ECE,MRCE,Hyderabad,mail:-vajjaparamesh12@gmail.com

Abstract—As the emerging field of machine learning, deep learning shows excellent ability in solving complex learning problems. However, the size of the networks becomes increasingly large scale due to the demands of the practical applications, which poses significant challenge to construct a high performance implementations of deep learning neural networks. In order to improve the performance as well to maintain the low power cost, in this paper we design DLAU, which is a scalable accelerator architecture for large-scale deep learning networks using FPGA as the hardware prototype. The DLAU accelerator employs three pipelined processing units to improve the throughput and utilizes tile techniques to explore locality for deep learning applications. Experimental results on the state-of-the-art Xilinx FPGA board demonstrate that the DLAU accelerator is able to achieve up to 36.1x speedup comparing to the Intel Core2 processors, with the power consumption at 234mW.

Index Terms—FPGA; Deep Learning; neural network; hard-ware accelerator.

I. INTRODUCTION

IN the past few years, machine learning has become perva-sive in various research fields and commercial applications, and achieved satisfactory products. The emergence of deep learning speeded up the development of machine learning and artificial intelligence. Consequently, deep learning has become a research hot spot in research organizations [1].

In general, deep learning uses a multi-layer neural network model to extract high-level features which are a combination of low-level abstractions to find the distributed data features, in order to solve complex problems in machine learning. Currently the most widely used neural models of deep learning are Deep Neural Networks (DNNs) [2] and Convolution Neural Net-works (CNNs) [3], which have been proved to have excellent capability in solving picture recognition, voice recognition and
other complex machine learning tasks.

However, with the increasing accuracy requirements and complexity for the practical applications, the size of the neural networks becomes explosively large scale, such as the Baidu Brain with 100 Billion neuronal connections, and the Google cat-recognizing system with 1 Billion neuronal connections. The explosive volume of data makes the data centers quite power consuming. In particular, the electricity consumption of data centers in U.S. are projected to increase to roughly 140 billion kilowatt-hours annually by 2020 [4]. Therefore, it poses significant challenges to implement high performance deep learning networks with low power cost, especially for large-scale deep learning neural network models. So far, the state-of-the-art means for accelerating deep learning algorithms are Field-Programmable Gate Array (FPGA), Application Spe-cific Integrated Circuit (ASIC), and Graphic Processing Unit (GPU). Compared with GPU acceleration, hardware accel-erators like FPGA and ASIC can achieve at least moderate performance with lower power consumption. However, both FPGA and ASIC have relatively limited computing resources, memory, and I/O bandwidths, therefore it is challenging to develop complex and massive deep neural networks using hardware accelerators. For ASIC, it has a longer development cycle and the flexibility is not satisfying. Chen et al presents a ubiquitous machine-learning hardware accelerator called DianNao [6], which initiated the field of deep learning pro-cessor. It opens a new paradigm to machine learning hardware accelerators focusing on neural networks. But DianNao is not implemented using reconfigurable hardware like FPGA, therefore it cannot adapt to different application demands. Currently around FPGA acceleration researches, Ly and Chow
[5] designed FPGA based solutions to accelerate the Re-stricted Boltzmann Machine (RBM). They created dedicated hardware processing cores which are optimized for the RBM algorithm. Similarly Kim et al [7] also developed a FPGA based accelerator for the restricted Boltzmann machine. They use multiple RBM processing modules in parallel, with each module responsible for a relatively small number of nodes. Other similar works also present FPGA based neural network accelerators [9]. Qi et al. present a FPGA based accelerator [8], but it cannot accommodate changing network size and network topologies. To sum up, these studies focus on implementing a particular deep learning algorithm efficiently, but how to increase the size of the neural networks with scalable and flexible hardware architecture has not been properly solved.

To tackle these problems, we present a scalable deep learning accelerator unit named DLAU to speed up the kernel computational parts of deep learning algorithms. In particular, we utilize the tile techniques, FIFO buffers, and pipelines to minimize memory transfer operations, and reuse the comput-ing units to implement the large-size neural networks. This approach distinguishes itself from previous literatures with following contributions:

1. In order to explore the locality of the deep learning

TABLE I
PROFILING OF HOT SPOTS OF DNN

Algorithms	Matrix Multiplication	Activation	Vector
Feedforward	98.60%	1.40%	
RBM	98.20%	1.48%	0.30%
BP	99.10%	0.42%	0.48%

application, we employ tile techniques to partition the large scale input data. The DLAU architecture can be configured to operate different sizes of tile data to leverage the trade-offs between speedup and hardware costs. Consequently the FPGA based accelerator is more scalable to accommodate different machine learning applications.

2. The DLAU accelerator is composed of three fully pipelined processing units, including TMMU, PSAU, and AFAU. Different network topologies such as CNN, DNN, or even emerging neural networks can be composed from these basic modules. Consequently the scalability of FPGA based accelerator is higher than ASIC based accelerator.

II. TILE TECHNIQUES AND HOT SPOT PROFILING

Restricted Boltzmann Machines (RBMs) have been widely used to efficiently train each layer of a deep network. Normally a deep neural network is composed of one input layer, several hidden layers and one classifier layer. The units in adja-cent layers are all-to-all weighted connected. The prediction process contains feedforward computation from given input neurons to the output neurons with the current network config-urations. Training process includes pre-training which locally tune the connection weights between the units in adjacent layers, and global training which globally tune the connection weights with Back Propagation process.

The large-scale deep neural networks include iterative com-putations which have few conditional branch operations, there-fore they are suitable for parallel optimization in hardware. In this paper we first explore the hot spot using the profiler. Results in Fig. I illustrates the percentage of running time including Matrix Multiplication (MM), Activation, and Vector operations. For the representative three key operations: feed forward, Restricted Boltzmann Machine (RBM), and back propagation (BP), matrix multiplication play a significant role of the overall execution. In particular, it takes 98.6%, 98.2%, and 99.1% of the feed forward, RBM, and BP operations. In comparison, the activation function only takes 1.40%, 1.48%, and 0.42% of the three operations. Experimental results on profiling demonstrate that the design and implementation of MM accelerators is able to improve the overall speedup of the system significantly.

However, considerable memory bandwidth and computing resources are needed to support the parallel processing, con-sequently it poses a significant challenge to FPGA implemen-tations compared with GPU and CPU optimization measures. In order to tackle the problem, in this paper we employ tile techniques to partition the massive input data set into tiled subsets. Each designed hardware accelerator is able to buffer the tiled subset of data for processing. In order to support the large-scale neural networks, the accelerator architecture are reused. Moreover, the data access for each tiled subset can run in parallel to the computation of the hardware accelerators.

Algorithm 1 Pseudocode Code of the Tiled

```
Inputs Require:
  Ni: the number of the input neurons
  No: the number of the output neurons
  Tile_Size: the tile size of the input data
  batchsize: the batch size of the input
  data for n = 0; n < batchsize; n + + do
    for k = 0; k < Ni; k+ = T ile Size do
      for j = 0; j < No; j + + do
        y[n][j] = 0;
        for i = k; i < k + T ile Size&&i < Ni; i + + do
          y[n][j]+ = w[i][j] x[n][i]
          if i == Ni       1 then
            y[n][j] = f(y[n][j]);
          end if
        end for
      end for
    end for
  end for
```

In particular, for each iteration, output neurons are reused as the input neurons in next iteration. To generate the output neurons for each iteration, we need to multiply the input neurons by each column in weights matrix. As illustrated in Algorithm 1, the input data are partitioned into tiles and then multiplied by the corresponding weights. Thereafter the calculated part sum are accumulated to get the result. Besides the input/output neurons, we also divided the weight matrix into tiles corresponding to the tile size. As a consequence, the hardware cost of the accelerator only depends on the tile size, which saves significant number of hardware resources. The tiled technique is able to solve the problem by imple-menting large networks with limited hardware. Moreover, the pipelined hardware implementation is another advantage of FPGA technology compared to GPU architecture, which uses massive parallel SIMD architectures to improve the overall performance and throughput. According to the profiling results depicted in Table I, during the prediction process and the training process in deep learning algorithms, the common but important computational parts are matrix multiplication and activation functions, consequently in this paper we implement the specialized accelerator to speed up the matrix multiplica-tion and activation functions.

III. DLAU ARCHITECTURE AND EXECUTION MODEL

Fig. 1 describes the DLAU system architecture which contains an embedded processor, a DDR3 memory controller, a DMA module, and the DLAU accelerator. The embedded processor is responsible for providing programming interface to the users and communicating with DLAU via JTAG-UART. In particular it transfers the input data and the weight matrix to internal BRAM blocks, activates the DLAU accelerator, and returns the results to the user after execution. The DLAU is integrated as a standalone unit which is flexible and adaptive

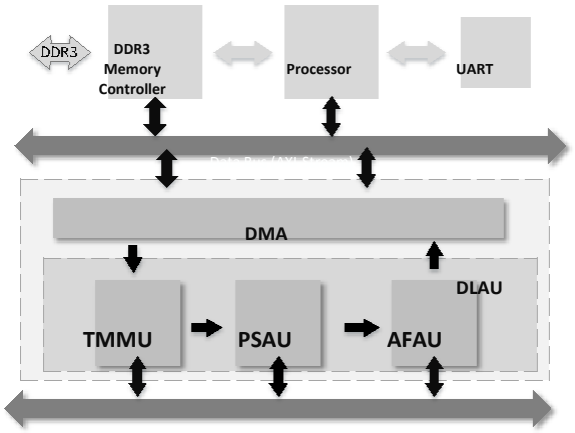


Fig. 1. DLAU Accelerator Architecture.

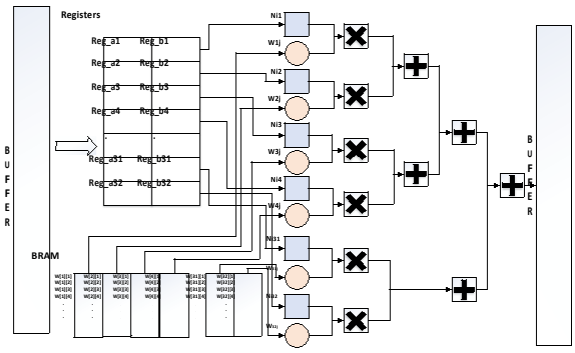


Fig. 2. TMMU Schematic Diagram.

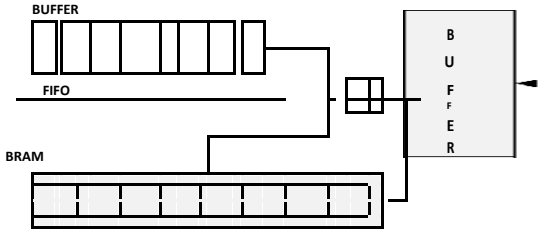


Fig. 3. PSAU Schematic Diagram

to accommodate different applications with configurations. The DLAU consists of 3 processing units organized in a pipeline manner: Tiled Matrix Multiplication Unit (TMMU), Part Sum Accumula2tion Unit (PSAU), and Activation Function Acceleration Unit (AFAU). For execution, DLAU reads the tiled data from the memory by DMA, computes with all the three processing units in turn, and then writes the results back to the memory.

In particular, the DLAU accelerator architecture has follow-ing key features:

FIFO Buffer: Each processing unit in DLAU has an input buffer and an output buffer to receive or send the data in FIFO. These buffers are employed to prevent the data loss caused by the inconsistent throughput between each processing unit.

Tiled Techniques: Different machine learning applications may require specific neural net-work sizes. The tile technique is employed to divide the large volume of data into small tiles that can be cached on chip, therefore the accelerator can be adopted to different neural network size. Consequently the FPGA based accelerator is more scalable to accommodate different machine learning applications.

Pipeline Accelerator: We use stream-like data passing mechanism (e.g. AXI-Stream for demonstration) to transfer data between the adjacent processing units, therefore TMMU, PSAU, and AFAU can compute in streaming-like manner. Of these three computational modules, TMMU is the primary computational unit, which reads the total weights and tiled nodes data through DMA, performs the calculations, and then transfers the intermediate Part Sum results to PSAU. PSAU collects Part Sums and performs accumulation. When the accumulation is completed, results will be passed to AFAU. AFAU performs the activation function using piecewise linear interpolation methods. In the rest of this section, we will detail the implementation of these three processing units respectively.

A. TMMU architecture

Tiled Matrix Multiplication Unit (TMMU) is in charge of multiplication and accumulation operations. TMMU is spe-cially designed to exploit the data locality of the weights and is responsible for calculating the Part Sums. TMMU employs an input FIFO buffer which receives the data transferred from DMA and an output FIFO buffer to send Part Sums to PSAU. Fig. 2 illustrates the TMMU schematic diagram, in which we set tile size=32 as an example. TMMU firstly reads the weight matrix data from input buffer into different BRAMs in 32 by the row number of the weight matrix (n=i%32where n refers to the number of BRAM, and i is the row number of weight matrix). Then, TMMU begins to buffer the tiled node data. In the first time, TMMU reads the tiled 32 values to registers Reg a and starts execution. In parallel to the computation at every cycle, TMMU reads the next node from input buffer and saves to the registers Reg b. Consequently the registers Reg a and Reg_b can be used alternately.

For the calculation, we use pipelined binary adder tree structure to optimize the performance. As depicted in Fig. 2, the weight data and the node data are saved in BRAMs and registers. The pipeline takes advantage of time-sharing the coarse-grained accelerators. As a consequence, this im-plementation enables the TMMU unit to produce a Part Sum result every clock cycle.

B. PSAU architecture

Part Sum Accumulation Unit (PSAU) is responsible for the accumulation operation. Fig. 3 presents the PSAU architecture, which accumulates the part sum produced by TMMU. If the Part Sum is the final result, PSAU will write the value to output buffer and send results to AFAU in a pipeline manner. PSAU can accumulate one Part Sum every clock cycle, therefore the throughput of PSAU accumulation matches the generation of the Part Sum in TMMU.

C. AFAU architecture

Finally, Activation Function Acceleration Unit (AFAU) implements the activation function using piecewise linear interpolation (y=ai*x+bi, x2[x_1,x_{i+1})). This method has been widely applied to implement activation functions with negligible accuracy loss when the interval between x_i and x_{i+1} is insignificant. Eq. (1) shows the implementation of sigmoid function. For x>8 and x -8, the results are sufficiently close to the bounds of 1 and 0, respectively. For the cases in -8<x 0 and 0<x 8, different functions are configured. In total we divide the sigmoid function into four segments.

$$f(x) = \begin{cases} 1 + a[b^{x}_{k} c_{x}]x \quad b[b^{x}_{kx} \quad c] & \text{if } 8 < x \quad 0 \\ > \quad 0 & \text{if } x \quad 8 \\ < \quad b\_c \quad b\_ \ c_{1} & \text{if } x > 8 \\ : \end{cases} \qquad (1)$$

Similar to PSAU, AFAU also has both input buffer and output buffer to maintain the throughput with other processing units. In particular, we use two separate BRAMs to store the values of a and b. The computation of AFAU is pipelined to operate sigmoid function every clock cycle. As a consequence, all the three processing units are fully pipelined to ensure the peak throughput of the DLAU accelerator architecture.

IV. Experiments and Data Analysis

In order to evaluate the performance and cost of the DLAU accelerator, we have implemented the hardware prototype on the Xilinx Zynq Zedboard development board, which equips ARM Cortex-A9 processors clocked at 667MHz and pro-grammable fabrics. For benchmarks, we use the Mnist data set to train the 784 M N 10 Deep Neural Networks in Matlab, and use M N layers weights and nodes value for the input data of DLAU. For comparison, we use Intel Core2 processor clocked at 2.3GHz as the baseline.

In the experiment we use Tile size=32 considering the hardware resources integrated in the Zedboard development board. The DLAU computes 32 hardware neurons with 32 weights every cycle. The clock of DLAU is 200MHz (one cycle takes 5ns). Three network sizes—64 64, 128 128, and 256 256 are tested.

A. Speedup Analysis

We present the speedup of DLAU and some other similar implementations of the deep learning algorithms in Table II. Experimental results demonstrate that the DLAU is able to achieve up to 36.1x speedup at 256 256 network size. In comparison, Ly&Chows work [5] and Kim et.als work [7] present the work only on Restricted Boltzmann Machine algorithms, while the DLAU is much more scalable and flexible. DianNao [6] reaches up to 117.87x speedup due to its high working frequency at 0.98GHz. Moreover, as DianNao is hardwired instead of implemented on a FPGA platform, therefore it cannot efficiently adapt to different neural network sizes.

Fig. 4 illustrates the speedup of DLAU at different network sizes-64 64, 128 128, and 256 256 respectively. Experi-mental results demonstrate a reasonable ascendant speedup

TABLE II
Comparisons between Similar Approaches

Work	Network	Clock	Speedup	Baseline
Ly&Chow [5]	256 256	100MHz	32	2.8GHz P4
Kim et.al [7]	256 256	200MHz	25	2.4GHz Core2
DianNao [6]	General	0.98GHz	117.87	2GHz SIMD
Zhang et.al [3]	256 256	100MHz	17.42	2.2GHz Xeon
DLAU	256 256	200MHz	36.1	2.3GHz Core2

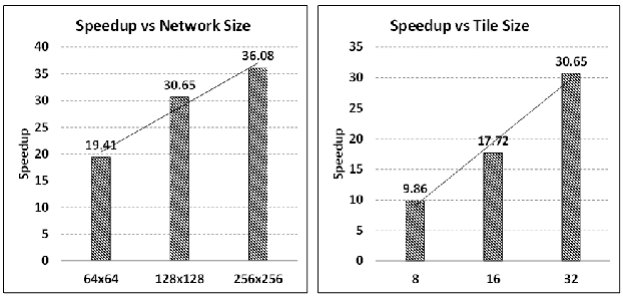


Fig. 4. Speedup at Different Network Sizes and Tile Sizes.

TABLE III
Resource Utilization of DLAU at 32 32 Tile Size

Component	BRAMs	DSPs	FFs	LUTs
TMMU	32	158	25356	32461
PSAU	1	2	754	632
AFAU	2	7	2216	3291
Total	35	167	28326	36384
Available	280	220	106400	53200
Utilization	12.5%	75.9%	26.6%	68.4%

with the growth of neural networks sizes. In particular, the speedup increases from 19.2x in 64 64 network size to 36.1x at the 256 256 network size. The right part of Fig. 4 illustrates how the tile size has an impact on the performance of the DLAU. It can be acknowledged that bigger tile size means more number of neurons to be computed concurrently. At the network size of 128 128, the speedup is 9.2x when the tile size is 8. When the tile size increases to 32, the speedup reaches 30.5x. Experimental results demonstrate that the DLAU framework is configurable and scalable with dif-ferent tile sizes. The speedup can be leveraged with hardware cost to achieve satisfying trade-offs.

B. Resource utilization and Power

Table III summarizes the resource utilization of DLAU in 32 32 tile size including the BRAM resources, DSPs, FFs, and LUTs. TMMU is much more complex than the rest two hardware modules therefore it consumes most hardware resources. Taking the limited number of hardware logic resources provided by Xilinx XC7Z020 FPGA chip, the overall utilization is reasonable. The DLAU utilizes 167 DSP blocks due to the use of the Floating-point addition and the Floating-point multiplication operations.

Table IV compares the resource utilization of DLAU with other two FPGA based literatures. Experimental results depict

TABLE IV
RESOURCE COMPARISONS BETWEEN SIMILAR APPROACHES

Implementation	FPGA	BRAMs	DSPs	FFs	LUTs
Ly&Chow [5]	XC2VP70	257	N/A	30403	29885
Kim et.al [7]	N/A	589824	18	11790	7662
DLAU	XC7Z020	35	167	28326	36384

TABLE V
POWER CONSUMPTION OF THE UNITS

Component	Power	Component	Power
Accelerator-TMMU	189mW	Processor	1307mW
Accelerator-PSAU	5mW	DDR Controller	177mW
Accelerator-AFAU	25mW	Peripherals	26mW
Accelerator-DMA	15mW	Clocks	70mW
Accelerator-Total	234mW	System Total	1814mW

that our DLAU accelerator occupies similar number of FFs and LUTs to Ly&Chow's work [5], while it only consumes 35/257=13.6% on the BRAMs. Comparing to the Kim et.al's work [7], the BRAM utilization of DLAU is insignificant. This is due to the tile techniques so that large scale neural networks can be divided into small tiles, therefore the scalability and flexibility of the architecture is significantly improved.

In order to evaluate the power consumption of accelerator, we use Xilinx Vivado tool set to achieve power cost of each processing unit in DLAU and the DMA module. The results in Table IV-B depict that the total power of DLAU is only 234mW, which is much lower than that of DianNao (485mW). The results demonstrate that the DLAU is quite energy efficient as well as highly scalable compared to other accelerating techniques. To compare the energy and power between FPGA based accelerator and GPU based accelera-tors, we also implement a prototype using the state-of-the-art NVIDIA Tesla K40c as the baseline. K40c has 2880 stream cores working at peak frequency 875MHz, and the Max Memory Bandwidth is 288 (GB/sec). In comparison, we only employ 1 DLAU on the FPGA board working at 100MHz. In order to evaluate the speedup of the accelerators in a real deep learning applications, we use DNN to model 3 benchmarks, including Caltech101, Cifar-10, and MNIST, respectively. Fig. 5 illustrates the comparison between FPGA based GPU+cuBLAS implementations. It reveals that the power consumption of GPU based accelerator is 364 times higher than FPGA based accelerators. Regarding the total energy consumption, the FPGA based accelerator is 10x more energy efficient than GPU, and 4.2x than GPU+cuBLAS optimizations.

Finally Fig. 6 illustrates the floor plan of the FPGA chip. The left corner depicts the ARM processor which is hard-wired in the FPGA chip. Other modules, including different components of the DLAU accelerator, the DMA, and memory interconnect, are presented in different colors. Regarding the programming logic devices, TMMU takes most of the areas as it utilizes a significant number of LUTs and FFs.

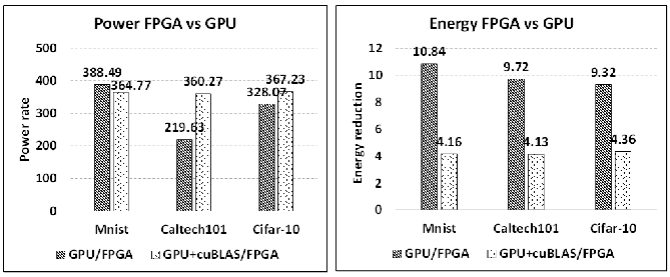


Fig. 5. Power and Energy Comparison between FPGA and GPU

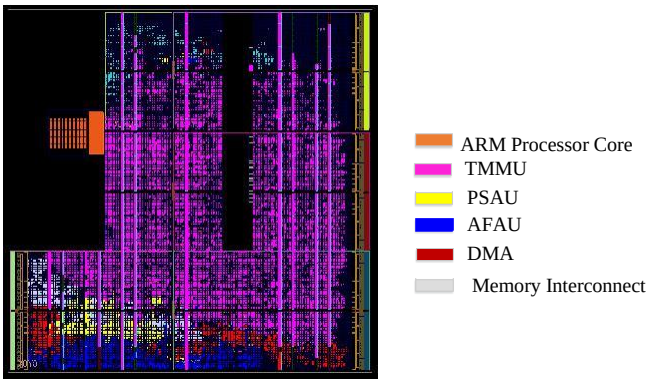


Fig. 6. Floorplan of the FPGA Chip

V. CONCLUSION AND FUTURE WORK

In this article we have presented DLAU, which is a scalable and flexible deep learning accelerator based on FPGA. The DLAU includes three pipelined processing units, which can be reused for large scale neural networks. DLAU uses tile techniques to partition the input node data into smaller sets and compute repeatedly by time-sharing the arithmetic logic. Experimental results on Xilinx FPGA prototype show that DLAU can achieve 36.1x speedup with reasonable hardware cost and low power utilization.

The results are promising but there are still some future directions, including optimization of the weight matrix and memory access. Also the trade-off analysis between FPGA and GPU accelerators is another promising direction for large scale neural networks accelerations.

REFERENCES

[1] LeCun, Y., Y. Bengio, and G. Hinton, Deep learning. Nature, 2015. 521: p. 436-444.
[2] Hauswald, J., et al. DjiNN and Tonic: DNN as a service and its implications for future warehouse scale computers. in ISCA 2015.
[3] Zhang, C., et al. Optimizing FPGA-based Accelerator Design for Deep Convolutional Neural Networks. in FPGA 2015.
[4] Thibodeau, P. Data centers are the new polluters. 2014 [cited 2015.
[5] Ly, D.L. and P. Chow, A high-performance FPGA architecture for restricted boltzmann machines, in FPGA 2009.
[6] Chen, T., et al., DianNao: a small-footprint high-throughput accelerator for ubiquitous machine-learning, in ASPLOS 2014. p. 269-284.
[7] Kim, S.K., et al. A highly scalable restricted boltzmann machine FPGA implementation. in FPL 2009.
[8] Qi Yu, et al. A Deep Learning Prediction Process Accelerator Based FPGA. CCGRID 2015: 1159-1162
[9] Jiantao Qiu, et al. Going Deeper with Embedded FPGA Platform for Convolutional Neural Network. FPGA 2016: 26-35

Division Circuit Using Reversible Logic Gates

RATHNAVATH PRIYADARSHINI
M.tech scholar , Dept of ECE,
Hyderabad,Priyadarshini.r73@gmail.com.

Abstract—In the recent years, reversible approachis becoming widely used in many domains, such as quantum computing, optical computing and ultra-low power VLSI circuit. Division has its application in the design of reversible Arithmetic Logic Unit (ALU). In this paper, we have exhibited a novel design of division sequential circuit using reversible logic gates. The proposed design of division block is based on reversible gates with reduction of garbage outputs, constant inputs, quantum cost and hardware complexity. The comparative results demonstrate that the proposed solution have less performanceand significantly better scalability than the currentdesigns

Keywords— Low power; reversible ALU; reversible gates; quantum cost; garbage outputs; hardware complexity; constant inputs; reversible division

I. INTRODUCTION

Power consumption is an essential issue in modern day VLSI (Very-Large-Scale Integration) designs. Consumption has become the essential limit of the increasing number of transistors per chip. In the recent years, many techniques have been developed at various levels in CMOS VLSI design.

- At system level and algorithmic level, techniques such as use of dynamic voltage Scaling (DVS) [1] and dynamic power management (DPM), as well as alternate encoding methods.[2].
- At architecture level, techniques such as use of parallel structures, pipelining [3], the clock gating [4].
- At circuit level, techniques such as use of the glitch [5].
- At the technology level, techniques [6] such as VT reduction, multi-threshold voltages have beenused.

Front of these diversity methods that are dedicated to improving the power VLSI design. In this paper we have encountered the reversible logic method [7]. It has wide applications in several technologies such as the nanotechnology, low power CMOS, optical information processing, bio information and DNA computing. In 1961 Rolf Landauer [8] showed that the irreversible logic gatesor the conventional logic gates, set-to-one like AND, OR, XOR and others dissipate a certain amount of energy caused the loss for each bit information during computation (at least kTln2 for each bit of information lost, where K is the Boltzmann's constant and T is the operating temperature).

In 1973, Bennett showed that [9] in order to avoid KT*ln2 joules of energy dissipation in a irreversible circuit, it must be built using reversible logic gates only, since there is no information loss happens in reversible circuits. In fact, division module is very important in the digital processing. Division process is one of the most difficult operations in the ALU and plays a big role in digital signal processing systems. This paper focuses on the design of division circuit sequential using reversible logic gates for positiveintegers.

II. BACKGROUND

In this section, we have exhibited the essential principles definitions to reversible logic along with the analysis of quantum cost.

A. ReversibleFunctions

Reversible function can be realized by reversible logic, is a bijection. Then there is a unique one-to-one mapping between an n-input vector and a corresponding n-output vector. Avoid leading output signals of gates to more than one input (Fan_out).Agatewithkinputsandkoutputsiscalledk*k gate. An k*k reversible logic gate can be denoted as $I_K - O_K$, where, $I_K = (I_1, I_2, \ldots, I_K)$ is the input vector and $O_K = (O_1, O_2, \ldots, O_K)$ is the output vector as show in fig.1 [10]. Reversible logic circuits have theoretically zero internal power dissipation because they do not lose data.

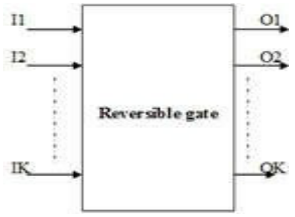


Fig.1. Reversible logic gate

B. Garbageoutputs

That defines the outputs of the reversible gate that is not used for further computations [11] [12]. The unutilized outputs from a gate are called ”garbage output” as show infig.2.

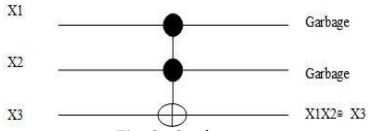


Fig.2. Garbage output

C. Constant inputs

Constant inputs (CIs) refer to the inputs which are used as control inputs and connected to logical '0' or '1' in order to obtain the logical function.
[13].

D. Quantum Cost

Quantum cost is the number of elementary 1*1 like NOT and 2*2 quantum logic gateslikeControlled ,Controlled and CNOT gates[14] [15] needed to realize the circuit.

E. Hardwarecomplexity

This refers to the total number of the logical calculations in a circuit that is measured by counting the number of AND operations, number of NOT operations and number of EX_OR operations [16]. To compute the hardware complexity of the reversible circuits we assume that:
a = Number of EX-OR gates. þ = Number of ANDgates.
6 = Number of NOTgates.

F. Synthesis of reversiblelogic

The design of reversible circuits significantly differs from the design of classical circuits. A reversible circuit should be designed using minimum number of reversible gates. One key requirement to achieve optimization is that the designed circuit must generate minimum number of garbage outputs. Equally they must use minimum number of constant inputs[17].
The fundamental rules for efficient reversible logic synthesis are:

- Reduction the number of quantum cost, garbage outputs, constant inputs and hardwarecomplexity.
- Avoid leading output signals of gates to more than one input (fan_out is notpermitted).
- Use a less number of reversible gates as conceivable to attain thegoal.

III.

BASIC REVERSIBLEGATES

Various reversible gates have been previously proposed by researchers/designers till now [24-30]. Each gate has a cost associated with it called the quantum cost. The NOT gate is a 1-q-bit gates and it has a quantum cost of zero. The N-bit Controlled-Gate has quantum cost of n-1. The Feynman gate can be operates as a controlled NOT (CNOT). If A is set to ‘1’ then the gate behaves as a Not gate, else a buffer gate. Feynman gate is widely used to surmount the fan-out problem as fan-out is not allowed in the reversible logic. It has a quantum cost of 1. The quantum cost of a Double Feynman gate is 2. The quantum cost of a Toffoli gate, TR gate and Peres gate are 4. Some examples of reversible logic gates are given by Table.1 [16-20]. These reversible gates help researchers/designers to design higher complex computing circuits. In this manuscript, we employ the reversible approach to realize a Division module.

TABLE1. Examples of some reversible logic gates

Gate	Quantum cost
Feynman gate	1
Not gate	1
Tr gate	4
Toffoli gate	5
Peres gate	4
Fredkin gate	5
BHA gate	4
BHC gate	5

IV. Relatedwork

A. Division operation

Essentially, the parameters of the division operation are dividend X and divisor Y as an input, and the quotient Q and remainder R as output, With X= Q*Y+R. For division, we use shift/adder division algorithm. At every step, we Shifted the register A and X (dividend) of 1 bit to the left. If the content of the register S.A is negative, then we add the contents of register A to Y. Else we subtract the contents of register A to
Y. If the result of subtraction/addition is positive (or zero) then, the quotient bit qi = 1. Else the quotient bit qi = 0. This process is repeated ntimes.

B. Components of division circuit

- Input reversible 8 bitMUX

Figure 3 shows the 2 inputs n-bit reversible MUX where Sel is the select input, and aretwo inputs. If Sel = 0, then or if S =1, then . This

reversible MUX consists of n the applicable criteria that follow.BHA[16]gateswhichisgeneratengarbageoutputsand needs 4n quantumcost

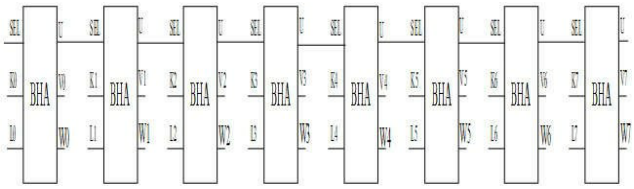
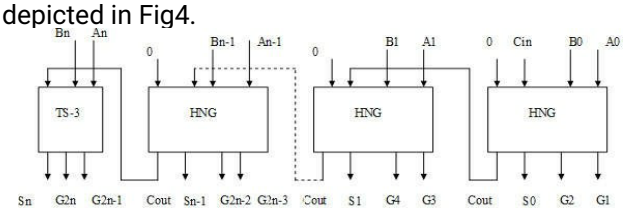


Fig 3. 2-input 8-bit reversible multiplexer circuit

- (8+1) BIT PARALLEL ADDER/SUBTRACTOR

To perform reversible realization of (n+1) adder/subtractor circuit we may use different combinations of any of the reversible logic gates. In [20] n bit reversible adder/ Subtractor using HNG gates is presented. The carry-out of the adder/ Subtractor is ignored in the proposed division circuit. Then, the implementation of (n+1) bit parallel adder/ subtractor requires n full Adder/Subtractor units and one TS-3 gate as depicted in Fig4.



II.

FIG 4. REVERSIBLE (N + 1) BIT PARALLEL ADDER WITH IGNORING OUTPUT CARRY

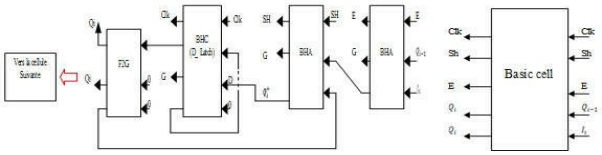
- N BIT REVERSIBLE PIPO LEFT-SHIFTREGISTERS

In n bit PIPO shift register as show in fig.15, during every clock pulse all information bits are loaded into the register. After shift operation all information bits are transferred together to their respective outputs by the same clock pulse. In [10], an n bits reversible PIPO right-shift register has been proposed. The researchers make it compatible for implementing left-shift register. Then, the basic elements used to design left shift register are multiplexer and D-latch. Fig 5 shows the proposed left shift register.

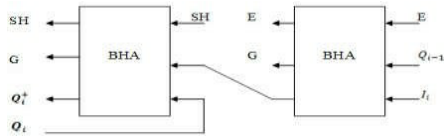
TABLE 2. Function table for reversible PIPO shift register.

SH	E	Finaloutput Q_i^+
0	0	Q_{i-1}(Leftshift)
0	1	I_i(Parallelload)
1		Q_i(Nochange)

- Basic cell for 1 bit reversible PIPO(Parallel Input-ParallelOutput)



- 3 to 1 reversibleMUX



- Basic cell for n bit reversible PIPO

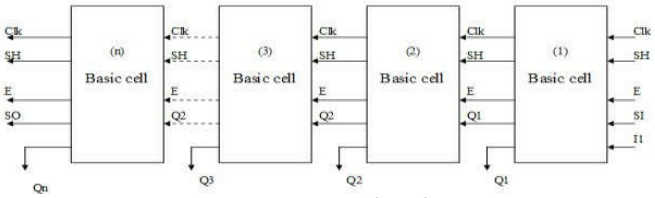


Fig 5. Reversible PIPO-left-shift register

- n- bit reversibleregister

The reversible D-Latch is used in [16] using BHC gate. Then, in order to implement n-bits reversible register we can use n BHC gates. Fig 6 shows the n-bit reversible register. It consists of n BHC gates, produces n garbage outputs, n constant inputs and needs 5n quantum cost. The hardware complexity is 4a+6þ+46.

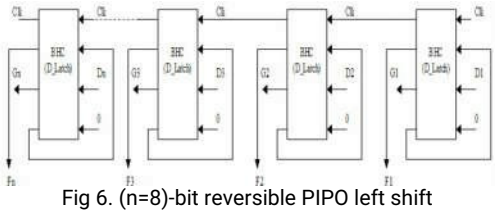


Fig 6. (n=8)-bit reversible PIPO left shift register.

- REVERSIBLE COUNTER (MOD-8)

A mod-8 counter reversible stores an integer value, and increments that value on each clock tick, and wraps around to 0 if the previous stored value was 7. The BHC gate [16] can be used as a D_ latch,where

$$Q_{(t)} = Clk.D + \overline{Clk}.Q_{(t-1)} \quad (1)$$

So we need three flip-flops D_

latch(BHC)where:D_0=1

$D_1 = Q_0$

$D_2 = Q_0.Q_1$

This reversible circuit produces a total number of 9 garbage outputs and 8 Constant inputs as show in Fig 7. It has a quantum cost of 28.

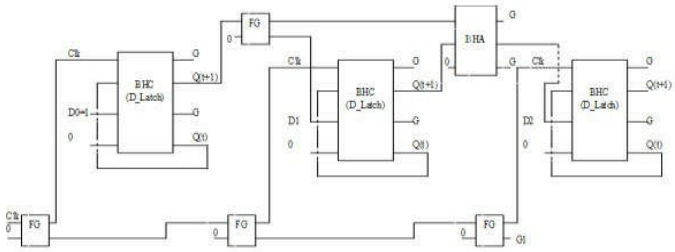
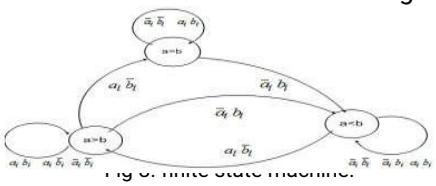


Fig 7. Reversible counter (mod-8)

- 4 Bit reversiblecomparator

In this section, two 4 bit numbers are compared with each other and the result shows that if one number is superior or fewer than other or if the two numbers are equal with each other. We use NOT gate, TR gate and BJN gate [17]. For example, assume A=A3 A2 A1 A0 and B=B3 B2 B1 B0, for comparing these two numbers, we use these finite state machine as show in Fig8.



Reversible comparator is demonstrated in Fig 9. This circuit produces a total number of 10 constant inputs, 15 garbage outputs, and 18 gates. The quantum cost of this comparator is 38.

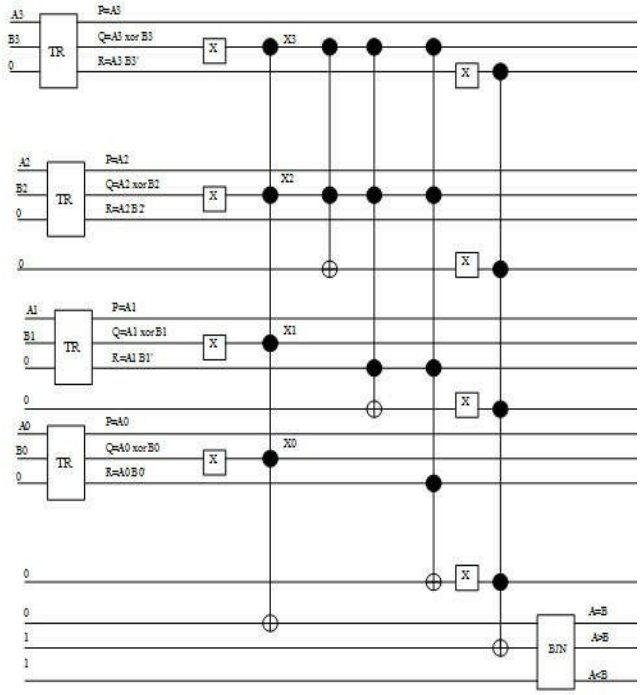


Fig 9. 4 bit reversible comparator.

Reversible n bitdivision

The reversible components are used to implement the division operation. In this paper an optimized reversible non restoring division using reversible logic gates is presented. This circuit includes (n=8).

- One n bit reversible PIPO left shiftregister
- One (n+1) bit reversible PIPO left shiftregister
- One 2 input reversible n bitMUX
- One 2 input reversible (n+1) bitMUX
- One (n+1) bit reversible parallel adder/Subtractor.

Algorithm:

Inputs:

S.A($S.A_{n-1}\ A_{n-2}\ldots\ A_0$) = 0 and

X($X_{n-1}\ X_{n-2}\ldots X_0$) =dividendand

Y($Y_{n-1}\ Y_{n-2}\ldots Y_0$)=divisor

Outputs:

Q($Q_{n-1}\ Q_{n-2}\ldots Q_0$)=quotientand

A($A_{n-1}\ A_{n-2}\ldots A_0$)=remainder.

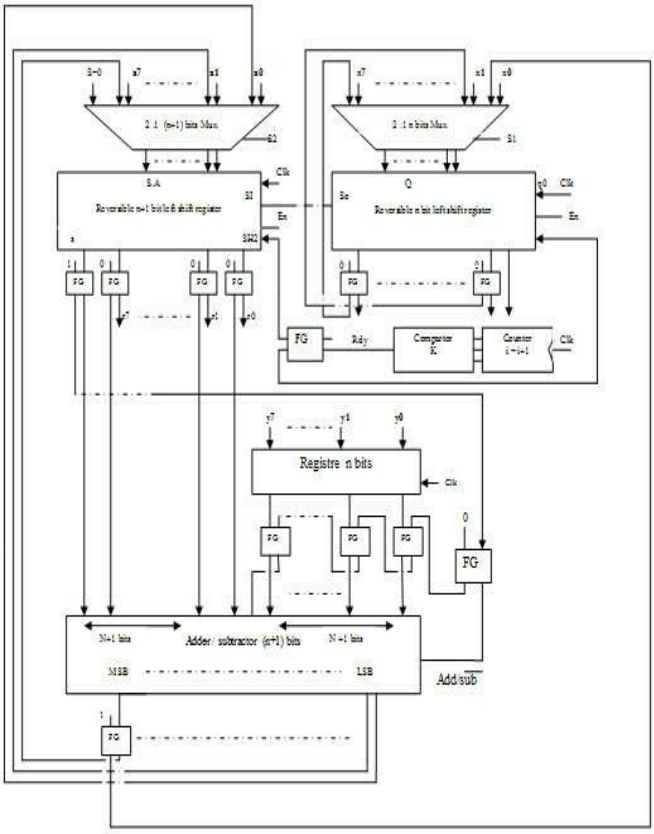


Fig 10. Proposed reversible 8 bit divider circuit

Fig 10 shows the proposed reversible divider circuit. It has 2 PIPO (parallel input-parallel output) left-shift registers. One is n+ 1 bits (n=8) named as S.A and other is n bits named as X. It equally contains an n bits register in order to store the divisor.Initially$S.A_{n-1}A_{n-2}.....A_0$=0,X($X_{n-1}X_{n-2}.....X_0$) = dividend, Y ($Y_{n-1}Y_{n-2}.....Y_0$) = divisor and rdy = 0. If the division process is completed then, the register X ($X_{n-1}X_{n-2}.....X_0$) contains the quotient and S.A ($S.A_{n-1}A_{n-2}.....A_0$) includes the remainder. On the other hand, when S1 = 0 then the two-input n-bit MUXselects dividend X ($X_{n-1}X_{n-2}.....X_0$), and if S2 = 0 then the two-input(n+1)bitMUXselects S=0andA($A_{n-1}A_{n-2}.....A_0$)= 0.Duringtheclockpulsewhen E=1,SH1=0andSH2=0, the output data from n bit MUX and (n+1)-bit MUX are loaded into X and S.A respectively. When E = 0, both S.A and X act as left-shift registers. Initially the value of S is not important, it is important just after the left shift of S.A&X (&=Concatenated), (S.A&X means SO of register X is connected to SI of S.A). If S is high ('1') then S.A-Y is performed, else, S.A+Y is computed. The complement of the MSB (most significant bit) is loaded into q0 bit position of register X. In addition, the sum is loaded into register S.A during next clock cycle (S1= 0). It includes 2n+1 clock signals to store the quotient value into register X. Finally, after 2n+1 clock signal, X stores the quotient and A stores the remainder indefinitely.

V.

D. DIVISION PROCESS EXAMPLE(14/2)

TABLE3. Division process example

Elementary operation	S.A3A2A1A0	X3X2X1X0
loading registers	00000	1110
Shift	00001	110-
Subtractor	00001 11101 1111 11	1100
Shift Adder	11111 000100 0001	100- 1001
Shift Subtractor ❖	00011 11101 1 00001	001- 0011
Shift Subtractor	00010 11101 1 00000 Final remainder	011- 0111 Quotient

Evaluation of the proposed reversible divider

Table 4 shows the characteristics of the proposed reversible division (n bits) without control circuit (counter, comparator).

TABLE 4. Characteristics of the proposed divider circuit.

Reversible components	Constant inputs	Quantum cost	Garbage outputs	Hardware Complexity HC (n=1)
n bit MUX 2 :1	0	4n	n	2a+ 4þ+36
(n+1) bit MUX 2 :1	0	4n+4	n+1	4a+ 8þ+66
(n)-bit register	n	5n	n+1	4a+ 6þ+46
n-bit left shift register	3n	15n	3n+2	10a+14þ+106
(n+1)-bit left shift register	3n+3	15n+15	3n+5	20a+28þ+206
(n+1)-bit parallel adder/ subtactor	n	6n+2	2n+2	7a+ 2þ
Feynman	2n+5	3n+3	0	6 a

- n-bit reversible MUX: number of garbage outputs = n, quantum cost = 4n, constant inputs=0 and the hardware complexity 2a+4þ+36.
- ❖ n+1-bitreversibleMUX:numberofgarbageoutputs = n+1, quantum cost = 4n+4, constant inputs=0 and the hardware complexity 4a+ 8þ+66.
- ❖ n-bit register: number of garbage outputs = n+1, quantum cost = 5n, constant inputs=n and the hardware complexity 4a+6þ+46.
- ❖ n-bit left shift register: number of garbage outputs = 3n+2, quantum cost = 15n, constant inputs=3n and the hardware complexity 10a+14þ+106.
- ❖ (n+1)-bit left shift register: number of garbage outputs = 3n+5, quantum cost = 15n+15, constant inputs=3n+3 and the hardware complexity 20a+ 28þ+206.
- ❖ (n+1)-bit parallel adder/ subtactor: number of garbage outputs = 2n+2, quantum cost = 6n+2, constant inputs=n and the hardware complexity 7a+ 2þ.
- ❖ Other gates (Feynman gate): number of garbage outputs = 0, quantum cost = 3n+3, constant inputs=2n+5 and the hardware complexity 7a+2þ.

VI. COMPARISON

In this section, we have compared the performance of the proposed design with some existing designs as depicted in table 4.

TABLE4.
Comparison

Conception	Constant inputs(n=8)	Quantum cost(n=8)	Garbage outputs (n=8)	Hardware complexity(n=1)
This work	88	440	99	53a + 62þ + 436
Existing design in [16]	89	454	103	65a +74 þ+526
Existing design in [10]	91	538	106	59a + 67þ + 336
Existing design in [13]	103	674	116	84a + 82þ + 426
Existing design in [14]	161	846	181	91a + 98þ + 506

Table 4 gives a comparison between the existing and the proposed 8 bits division designs. The proposed design reduces the garbage output, quantum cost, constant inputs as well as the hardwarecomplexity.

VII. CONCLUSION AND FUTUREWORK

In this manuscript, we proposed one approach for designing reversible division circuits. Then we have optimized our design. We have compared this proposed design with the existing designs. The proposed module for non-restoring division has better performance in terms of constant inputs, garbage output, and quantum cost as well as the hardware complexity than that of existing designs. In the future works we will design complete reversible ALU, develop reversible hardware description language and reversible synthesis tools.

REFERENCES

[1] Dongkun Shin, and Jihong Kim, "Intra-Task Voltage Scheduling on DVS-Enabled Hard Real-Time Systems", IEEE transactions on computer-aided design of integrated circuits and systems, VOL. 24, NO. 10, pp 1530-1549, October2005.

[2] Bishop Brock and Karthick Rajamani, "Dynamic Power Management for Embedded Systems". IEEE international SoC conference, pp 17-20, September 2003.

[3] Anantha P. chandrakasan and robert W. Brodersen, "Minimizing power consumption in digital CMOS circuits". IEEE, vol 4, pp 498-523, April 1995.

[4] Dushyant Kumar Soni1, Ashish Hiradha, "Dynamic Power reduction of synchronous digital design by using of efficient clock gating technique", International Journal of Engineering and Techniques Volume 1, pp 18- 23, May2015.

[5] Vikas, Deepak, "A review on glitch reduction techniques", International Journal of Research in Engineering and Technology. Volume: 03 pp 145-148, Feb-2014.

[6] Dhananjay E. Upasani, Sandip B. Shrote. "Standby Leakage Reduction in Nanoscale CMOS VLSI". International Journal of Computer Applications Volume 7– No.5, September2010.

[7] Onkar Nath Sabran, Shrikant Ahirwar, "An Extensive Review on Reversible Logic in VLSI", International journal of scientific progress and research. Vol-01,2014.

[8] R. Landauer, "Irreversibility and heat generation in the computing process". IBM Journal of Research and Development, vol. 5, pp. 183- 191,1961.

[9] C.H.Bernet, "Logical Reversibility of computation", IBM Journal of Research and Development, pp 525-532,1973.

[10] Noor Muhammed Nayeem, Md. Adnan Hossain, Md. Mutasimul Haque, Lafifa Jamal, Hafiz M. Hasan. Novel Reversible DivisionHardware,
IEEE, pp 1134-1138. 2009

[11] Rohini H Dr. Rajashekar S Dr. PriyatamKumar. " Design of basic sequential circuits using reversible logic". IEEE, pp 2110-2115.2016

[12] Deeptha A, Drishika Muthanna, Dhrithi M, Pratiksha M, B S Kariyappa, " Design and optimization of 8 bit ALU using reversible logic", IEEE, pp 1632-1636. 20-21May.2016.

[13] Faraz Dastan1 and Majid Haghparast," A novel nanometric fault tolerant reversible divider". International Journal of the physical Sciences Vol. 6(24), pp. 5671-5681, 16 October,2011.

[14] Faraz Dastan1 and Majid Haghparast. "A Novel Nanometric Reversible Signed Divider with Overflow Checking Capability". Research Journal of Applied Sciences, Engineering and Technology vol 4(6) pp 535-543, 2012

[15] Pallav Gupta Abhinav Agrawal, and Niraj K. Jha. " An algorithm for synthesis of reversible logic circuits" IEEE, pp 2317-2330. November 2006.

[16] A. Bolhassani and M. Haghparast, "Optimised reversible divider circuit", Int. J. Innovative Computing and Applications, pp 13-33,2016.

[17] Nagamani A, Jayashree H,H R Bhagyalakshmi, " Novel low power comparator design using reversible logic gates". IJCSE, pp 566-574, sep 2011

[18] A. Majumder, P.L. Singh, Efficient design and analysis of N-bit reversible shift registers. Procedia Computer Science, pp 199 – 208, 2015

[19] Deeptha A, Drishika Muthanna, Dhrithi M, Pratiksha M, BS Kariyappa , "Design and optimization of 8 bit ALU using reversible logic", IEEE, pp 1632-1636. May, 21, 2016

[20] Subramanian Saravanan, Ila Vennila, Sudha Mohanram, Design and Implementation of efficient reversible arithmetic and Logic Unit. SciRes, pp 630- 642.2016.

[21] Robert Wille, Anupam Chattopadhyay, Rolf Drechsler. From Reversible Logic to Quantum Circuits: Logic Design for an Emerging Technology, IEEE, pp 268-274.2016

Design of an Energy Efficient IoT enabled Smart System based on DALI network over MQTT protocol

[1]K. Premalatha

[1]PG Student, Dept. of ECE, MRCE, Hyderabad, India, Email: premalathakotthapeta124@gmail.com

ABSTRACT- The term "Internet of Things" or "IoT" refers to a hot and virgin area that is gaining importance day by day with increasing advancement in technology. With the help of micro-controllers like Arduino and micro- computers like Raspberry Pi, cheap devices can be used to measure sensor data and send it over the internet. Also, DALI (Digital Addressable Lighting Interface) is a new module of communication between the electrical equipment and a server or computer that will result in an effective realization of a smart-grid. This paper covers the communication of a DALI network of a group of lights with a Raspberry Pi (broker) over MQTT (Message Queuing Telemetry Transport) protocol keeping the concept of IoT in the background. This method proved to optimize electricity usage by optimal controlling the intensity of LEDs at various floors of the CDAC smartbuilding.

Key words – Internet of Things (IoT), DALI, MQTT, Paho C Client Library, Smart Grid, Raspberry Pi.

I. INTRODUCTION

In line with the mainstream technological innovations the emerging IoT technology aims at improving the quality of life of general masses and the efficiency of the civic amenities in a sustainable fashion while optimising economic investments by private and public sector. The IoT applications can be found in various vital fields such as: energy, health, transportation, environment, etc. Millions of application in all these fields can work optimally if there is a strong interconnection among IoT enabled devices [1]. Interconnection among IoT enabled devices via communication channels and protocols is not only a simple technological subject matter but it interests also other characteristics such as privacy, standardization, legal issues, etc. [2]. This unsurprisingly induces newer innovative challenges in IoT environment, which inspires industrial and academics researchers to go further in IoT sustainability research [3]. Confidently, the technological interfacing and communications among the IoT enabled devices (things) is essential to empower the IoT environment [4]. This is the reason why few vital functions will be performed by few key technological components and techniques, which have evolved to guarantee interconnection among heterogeneous devices adapting to the usage of very less supplies in terms of both computational time, memory, functionalities and energy resources.

This paper offers an analysis report of a small project done by the authors at CDAC, Pune to configure the electrical arrangement of a building in a smart manner using the concepts of DALI and IoT.

Accordingly, the rest of the paper is organized as follows: Section II identifies the basic apparatus required for the project and a brief description about the various protocols and modules that have been exhausted in the project. Section III provides an overview of the method through which the project was carried out and the working principles of the project. Section IV presents a brief of the results achieved from the project and an analysis of how smart grids are efficient investments for people in the time to come. Section V consists of the conclusion and outlook. Following the conclusion is a short note of acknowledgement by the authors to the people who have helped during the course of this project.

II. SYSTEM MODELLING

A. Hardware and Software used

In order to carry out this project, we made use of the following set of hardware:

- A Raspberry Pi 3 Model B
- DALI LED drivers, preferably LCM 60 DA.
- Group of LED lights

The software comprised of:

- Linux/Debian terminal

- Eclipse Paho client library for MQTT communication
- Any C programming IDE
- MQTT mosquitto broker

Before getting into the process of performing the project, it is imperative to acquaint ourselves with the basic concepts behind the modules and protocols in use:

1) MQTT: MQTT is a lightweight messaging protocol that provides resource-constrained network clients with a simple way to distribute telemetry information. The protocol, which uses a publish/subscribe communication pattern, is used for machine-to-machine (M2M) communication and plays an important role in the Internet of Things (IoT). MQTT allows devices to send (publish) information about a given topic to a server that functions as an MQTT message broker. The broker then pushes the information out to those clients that have previously subscribed to the client's topic. To a human, a topic looks like a hierarchical file path. MQTT is a good choice for wireless networks that experience varying levels of latency due to occasional bandwidth constraints or unreliable connections.

Should the connection from a subscribing client to the broker get broken, the broker will buffer messages and push them out to the subscriber when it is back online. Should the connection from the publishing client to the broker be disconnected without notice, the broker can close the connection and send subscribers a cached message with instructions from the publisher.

There are many platforms of MQTT protocol, the most popular being the HiveMQ and the Mosquitto platforms.

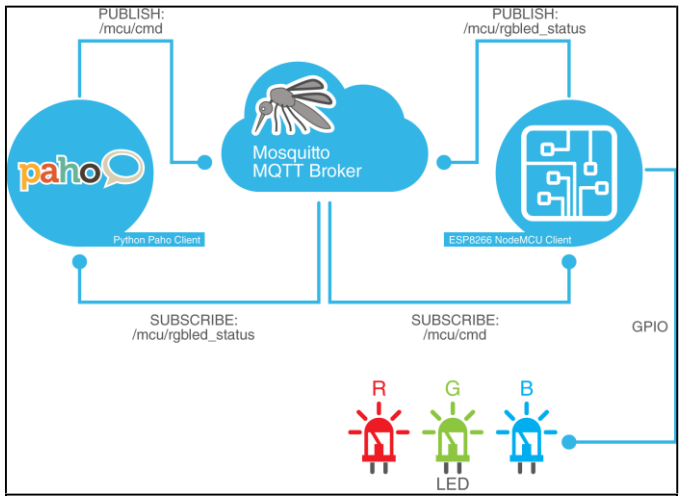


Fig. 1 Interaction of MQTT broker with the publisher and subscriber. [5]

In Fig.1, a Mosquitto MQTT broker acts as the server for communication between two clients (ESP8266 Node MCU and Paho Python), both of which can act as either a subscriber or a publisher. However, at a time, one will act as the publisher while the other will be the subscriber. This communication happens over a topic which is common to both ends. This maintains the singularity in communication over the same broker among various clients without any interference .

2) Eclipse Paho: The word ***pāho*** is a Maori word that means *to broadcast, make widely known, announce, disseminate, transmit*. The Paho project was created to provide scalable open-source implementations of open and standard messaging protocols aimed at new, exisiting, and emerging applications for Machine-to-Machine (M2M) and Internet of Things (IoT). Objectives include effective levels of decoupling between devices and applications, designed to keep markets open and encourage the rapid growth of scalable Web and Enterprise middleware and applications. Paho initially started with MQTT publish/subscribe client implementations for use on embedded platforms, and in the future will bring corresponding server support as determined by the community [6].

Paho client libraries can be written in many programming languages, viz. C, C++, Java, Python, etc. For this project, we performed Paho client programming in C.

3) DALI: DALI is a data protocol and transport mechanism that was jointly developed and specified by several manufacturers of lighting equipment. The common platform of DALI enables equipment from different manufacturers to be connected together. DALI network consists of a controller and one or more lighting devices (e.g., electrical ballasts, LED drivers and dimmers) that have DALI interfaces [10]. The controller can monitor and control each light by means of a bi-directional data exchange. The DALI protocol permits devices to be individually addressed and it also incorporates Group and Scene broadcast messages to simultaneously address multiple devices (e.g., "Group 1 goto 100%" or "Recall Scene 1"). DALI devices include LED drivers, fluorescent HF ballasts, low voltage transformers, PE cells, motion detectors, wall switches and gateways to other protocols. There can be up to 64 DALI devices on a single DALI network. Sites requiring more than 64 devices are implemented by having multiple separate DALI networks, each with up to 64 devices. These separate networks are then linked together with DALI gateways and a data bus running a high level protocol.

III. METHODOLOGY

This project works on the basic MQTT protocol, i.e. a publisher sends some data to an MQTT broker (created on the Raspberry Pi itself) which is further transferred to the subscriber. Both the publisher and subscriber are subscribed to the same topic. Each Raspberry Pi system is capable of handling at most 64 lights (clients) as per the DALI protocol.

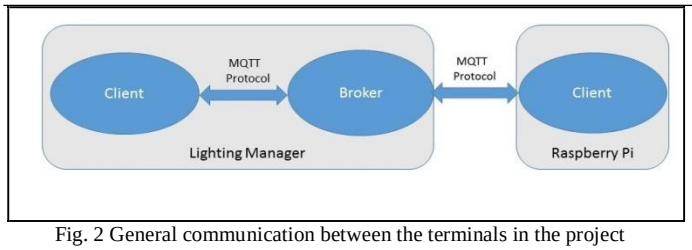
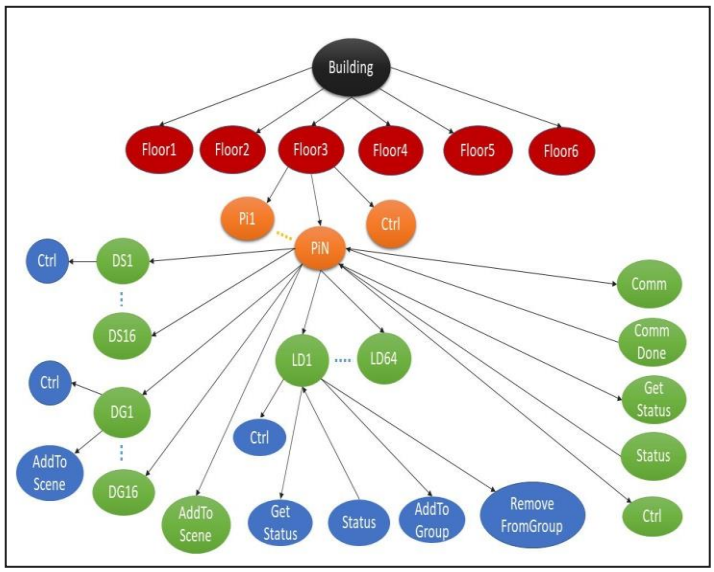


Fig. 2 General communication between the terminals in the project

Fig.2 gives a pictorial representation of the skeleton of the project where Lighting manager represents the console of DALI drivers, each of which connects to maximum LEDs



(nodes).

Fig. 3 Sequential range of topics undertaken by current research methodology for optimal energy saving with more than 64 devices DALI Network

The first stage in the project is to begin with the installation of Paho C client library on the Raspberry Pi. Thereafter, the mosquitto MQTT platform is installed. Once the installation process is over, the structure of topics is brought into existence. As is evident from Fig.3, the range of topics available is diverse. Each topic corresponds to a different set of topic. The hierarchy of topics is as follows:

a) *Building -> Floor n -> Pi n*: Each floor can have many Raspberry Pis to drive numerous DALI drivers. Hence, once the communication happens over this topic, the user has the option to control the state of either one or more than one LEDs lights on that floor through the corresponding Pi.

b) *Building -> Floor n -> Ctrl*: If the user wishes to control the entire floor and all Raspberry Pis and DALI drivers on that floor, this topic will be sent.

c) *Building -> Floor n -> Pi n -> Comm*: This topic is sent by the publisher to start the communication with the subscriber (DALI drivers).

d) *Building -> Floor n -> Pi n -> Comm Done*: This topic returns a callback signal to the broker whether successful communication has been formed with the client. If not, the broker will retry communication with the client until it's done.

e) *Building -> Floor n -> Pi n -> Get Status*: If the user wishes to know the current brightness status of the lights or of a particular DALI driver, this topic will be communicated with.

f) *Building -> Floor n -> Pi n -> Status*: The client returns its current status to the user as a return call to the above topic.

g) *Building -> Floor n -> Pi n -> Ctrl*: The user can get control over a single Pi on a particular floor and control all the DALI drivers with a single command through this topic.

h) *Building -> Floor n -> Pi n -> DS N -> Ctrl*: The term DS in this topic implies 'DALI Scene' and N represents the particular DALI driver that has 4 nodes. This topic enables the user to only control the brightness scene and make changes to it, perhaps like moderating the dim and bright lighting according to the situation. For instance, if it is a projection room, the lights are bright initially but may be automatically dimmed by the user once the projection of a presentation or movie begins in the room.

i) *Building -> Floor n -> Pi n -> DG N ->Ctrl*: The term DG here signifies 'DALI Group' and N is same as in the above topic. Here, there are total 16 DGs under each Pi and each DG consists of 4 nodes. We can control the settings and state of each group here, however cannot change the scene settings.

j) *Building -> Floor n -> Pi n -> DG N -> Add to Scene*: This topic is useful when the user wishes to access a particular DG and add some brightness settings to its current scene without altering it.

k) *Building -> Floor n -> Pi n -> Add to Scene*: Instead of changing the scenes of each DG or DS, one can also add settings to the scene of all 64 nodes under one Pi through this topic.

l) *Building -> Floor n -> Pi n -> LD N -> Ctrl*: The term LD in this topic refers to 'LED DALI', i.e. instead of having the nodes (LEDs) as groups, they can also be individually placed under a Raspberry Pi

thereby giving access to a Pi to 64 LDs. Hence, with this topic the state of the LDs can be controlled individually.

m) *Building -> Floor n -> Pi n -> LD N -> Get Status*: Same as topic e), only difference being that an individual LD's status is being sought for this time by the user.

n) *Building -> Floor n -> Pi n -> LD N -> Status*: Similar to topic f), this one returns the status of the LD which was sought by the user in the earlier topic.

o) *Building -> Floor n -> Pi n -> LD N -> Add to Group*: There is also an option to add the individual LDs to an existing DG or into a new one, keeping in mind that maximum number of nodes in each group should not exceed 4.

p) *Building -> Floor n -> Pi n -> LD N -> Remove from Group*: The LDs can be removed from the groups that they are a part of too.

How does it work?

The above range of topics describe the vast scope of available options to the user for controlling and monitoring a network of lights either in a home or even in a building. However, this is achieved through programming in Paho C client library.

We program both ends of the broker in such a way that firstly a client is created. A user-friendly environment is developed where the topics are listed on the screen, and the user can select a topic of his choice. Once the topic is also initialized, the subscriber client however tries connection with all the topics. Depending on the topic that the publisher has connected to, the subscriber reaches out to the set of topics corresponding to that topic and returns its status to the publisher via the broker.

Depending on the choice entered by the user, the corresponding topic gets generated on the publishing side. This choice of topic is sent to the MQTT broker (Raspberry Pi), which further channelizes the communication of data with the subscriber over this topic. The subscriber however generates all possible topics and tries to connect to each one of them at very short intervals after creating the client. The communication takes place when the topics of both publisher and subscriber match with each other. At this instant, the message from the publisher is published to the subscriber, and the resulting set of operations corresponding to that topic are performed. Eventually, we finally see the result physically with the varying brightness level of the lights or LEDs. The same process continues and various publishers can publish to the same subscriber over different topics. Also, the connection process is made in such a way that even if the connection is lost or broken from either sides, the client tries to reconnect with the broker and does not disconnect, unless disconnected manually by the user.

IV. RESULTS

Although we approached this project at its beginning stage, yet by implementing the proposed methodology many results were realized by us towards the end of it. The terminal windows consisting of the publishing and subscribing operations has been shown in section III of this paper.

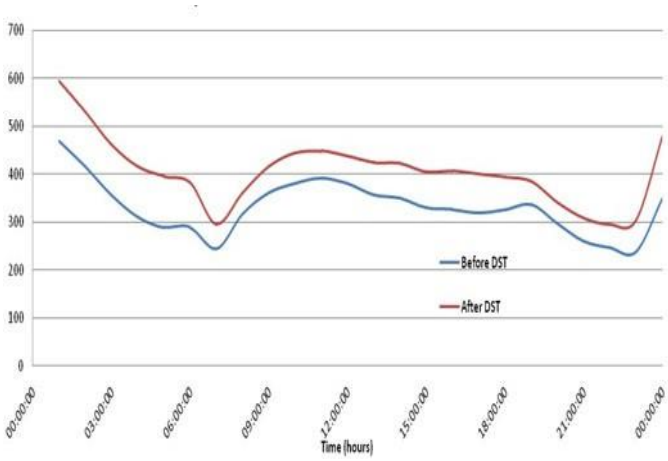


Fig. 4 Graphs of energy saving (in kWh) on Y-axis before and after DALI Scheme Transfer (DST) in a day

The analytics of this shows that on the peak periods of usage, the load can be increased and decreased vice versa via DALI controller. Also, the user desired topic can be sent to the MQTT broker (Raspberry Pi), which further publishes the data to the subscriber for optimally setting the brightness of LEDs as shown in Fig. 5a and 5b. This causes the energy saving in the electricity usage as shown in Fig. 4, and this will play a huge role in the energy saving in urban areas paving the way for a smarter demand response on consumer side.

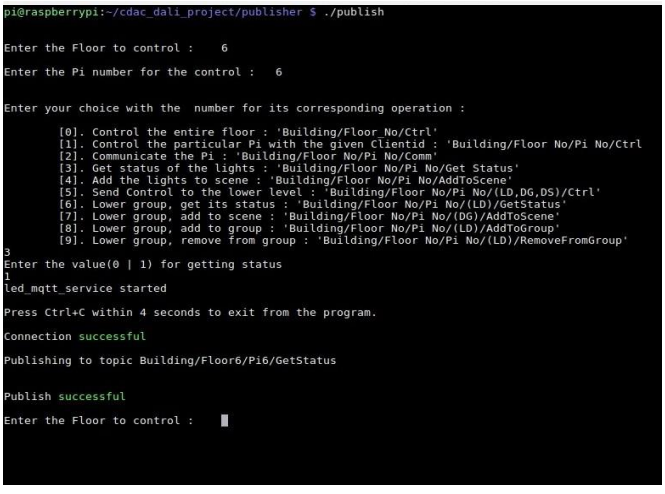


Fig. 5a Terminal window for the publisher

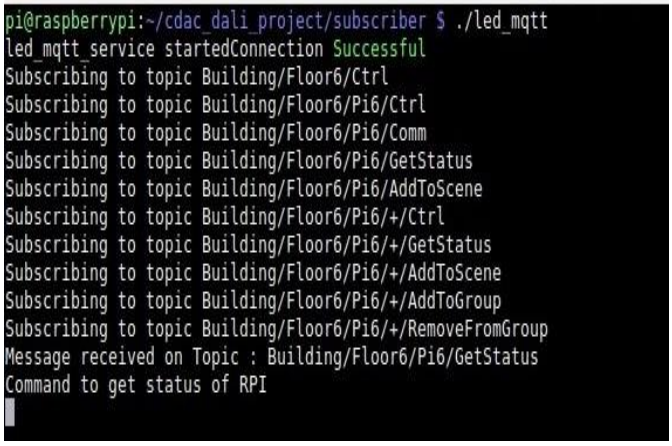


Fig. 5b. Terminal window for the subscriber

V. Conclusion

The coming decade is going to witness a revolutionary change in the manner of usage of energy and smart monitoring systems are going to come in vogue. We have been able to successfully implement a smart energy saving methodology via communication of a DALI network with a Raspberry Pi over MQTT (Message Queuing Telemetry Transport) protocol keeping the concept of IoT as the background for optimizing the brightness control of LEDs at different floors in the smart building of CDAC Pune. The constraint of using only 64 DALI devices on a single DALI network was overcome by having multiple separate DALI networks, each with up to 64 devices in the entire set up with reduced delay and maximum energy saving.

References

[1] aheeruddin and Munish Manas, “A New Approach for the Design and Development of Renewable Energy Management System through Microgrid Central Controller”, Energy Reports, Elsevier Inc., vol. 1, pp. 156-163, 2015.

[2] https://en.wikipedia.org/wiki/Digital_Addressable_Lighting_Interface

[3] Zaheeruddin; Manas, M., "Development of Self-Sustainable Technologies for Smart Grid in India," in Computational Intelligence & Communication Technology (CICT), 2015 IEEE International Conference on , vol., no., pp.563-568, 13-14 Feb. 2015.

[4] Zaheeruddin and Munish Manas “Analysis of Design of technologies, tariff Structures and regulatory policies for sustainable growth of the Smart grid” in Taylor and Francis’s “Energy Technology and Policy” Journal, Volume 2, issue 1, pp. 28-38, 2015.

[5] B. Saharia and Munish Manas, “Viability analysis of Photovoltaic/Wind hybrid standalone Distributed Generation based system as applicable in isolated community of north-east region of India”, Distributed Generation and Alternative Energy Journal , vol. 32, issue 1, PP. 49-80 • January 2017.

[6] W. Wu, M. Wu and Y. Liu, "A Design of Embedded DALI Controller," *2006 4th IEEE International Conference on Industrial Informatics*, Singapore, 2006, pp. 1237-1240.

[7] doi: 10.1109/INDIN.2006.275816

Verification and Simulation of New Designed NAND Flash Memory Controller

Akarapu Natraj Renu

PG Student, Dept. of ECE, MRCE, Hyderabad, India, Email: natrajrenu@gmail.com

Abstract— **In this paper a NAND flash memory controller was designed. For the best use of NAND type flash memory we design a new Arithmetical and Logical Unit (ALU) for calculating increment, addition, subtraction, decrement operations etc. In this memory controller we design single memory cell, memory module, a decoder etc. These all are encapsulated inside a controller and this is on top most in hierarchy. NAND flash memory is a non volatile storage media used in today daily life electronic equipments. NAND flash memory is programmed on page by page basis. Typically programming time is very less few micro second per page. This NAND flash memory controller architecture can be used with a real secure digital card, multimedia card (SD/MMC), digital cameras etc. The NAND Flash memory controller can be an internal device, built into the application processor or host, or designs can incorporate an external, stand-alone chip. Experimental results show that the designed controller give good performance and full fill all the system specifications. We have used FPGA chip for download our code.**

Keywords-Flash memory; Non Volatile; Arithmetical and Logical Unit (ALU); Encapsulation; Field Programmable Gate Array (FPGA);Secure Digital card/Multi Media Card (SD/MMC);Hard Disk Drive (HDD);

I. INTRODUCTION

Flash memory is being widely used as a storage medium in mobile devices because of its low power consumption, small form factor, and high resistance to shock and vibration. As the density of a flash memory chip increases and the price continues to drop, the flash memory is being adopted in more diverse storage applications. Flash memory is the combination of two technologies-EPROM and EEPROM. The term "Flash" means - A large chunk of memory (memory cell) could be erased at one time". On the other hand in EEPROM each byte is erased one by one manner.

Flash memory has characteristics that are different from conventional storage devices such as HDD. Thus, specialized hardware and software are required to use flash memory as a storage device. The role of flash memory software is particularly important because it has to deal with the peculiarities of flash memory. It also needs to be diverse because flash memory is used in a wide spectrum of applications ranging from micro-embedded systems (e.g., a sensor-node) to large-scale servers [5]. Flash memory controller offers higher capacity for fast data transfer and random access of memory in I/O operations [1]. It is possible to design a simple memory mapped interface to hardware with NAND flash memory. NAND flash memory controller has also a bidirectional bus in between peripheral devices and controller, controller internal bus for various functional blocks and controller with memory connections. For the improvement of product lifetime and system performance we always design an excellent NAND flash memory controller. Single Level Cell (SLC) and Multi Level Cell (MLC) two techniques are used for storing data in memory cells. SLC offers 100,000 erase program cycle while MLC offers about 10,000 erase program cycle [2].

NAND flash memory controller has also a bidirectional bus (Data, Address, and Control). NAND flash cell are placed together for saving 60% cell size over NOR flash cells. NAND flash memory controller provides a serial access of data blocks in a very high speed [6]. Single Level Cell (SLC) and Multi Level Cell (MLC) two techniques have been used for storing data in memory cells. A software called Flash Transaction layer (FTL) use for wear leveling and bad block management technique. All vendors provide FTL software [9].

The pipelining and parallel processing concepts are applied for systematic design approach of a systolic array processor. The systolic array architecture and iterative Very Large Scale Integration (VLSI) architecture is applied for good performance. It makes the circuit design easy for implementation [8].

II. NAND FLASH MEMROY CONTROLLER ARCHITECHTURE

The capacity of NAND flash device is improved day by day, architecture are also improved day by day. Latest overall structure of NAND flash device is looks very similar to its conventional structure. A NAND flash controller implements memory mapped interface [10]. A multiplane array packs contains its own set of Cache/Data registers, more memory cells on a die and partitioned it into several plans. In multiplane array packs all operations performs parallels. Inside a single flash array operation time multiple pages of data can be programmed, read, write, fetched etc. so average data access time is small. In multiplane commands some addresses are restricted.

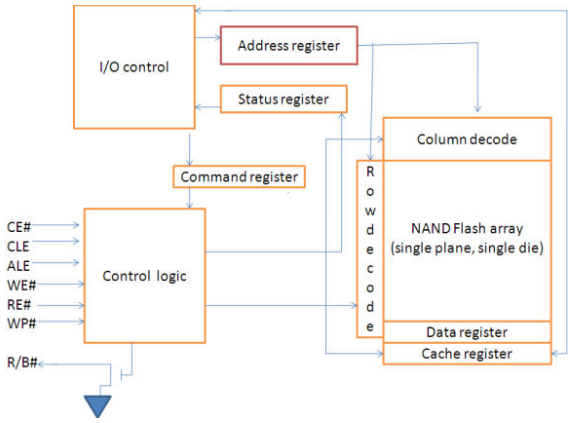


Figure 1. Block Diagram of a NAND Flash Memory Controller

NAND flash controller has flash chips. In today life it is widely used in storage devices. One more recent application for flash memory is as a replacement for hard disks. Flash memory does not have the mechanical limitations and latencies of hard drives, so a Solid State Drive (SSD) is attractive when considering speed, noise, power consumption, and reliability. It always follo ws Open NAND Flash Interface (ONFI) standard [3].

In NAND flash device we have mainly I/O control block, control logic, NAND flash array. A NAND flash array includes two dimensional NAN D flash cells, Row/Column address decoder and cache / data registers [11]. It has a shared, multiplexed, bidirectional (command, address and data) I/O bus. Figure-1 shows the block diagram of a NAND Flash Memory Controller. A multiplane array packs contains its own s et of Cache/Data registers, more memory cells on a die partitioned into several plane. In multiplane array packs, a ll operations are performed in parallel. Thus, inside a si ngle flash array, multiple pages of data can be programm ed, read, write, fetched, so average data access time is reduced. In multiplane commands some addresses are restricted. So new NAND flash chip with multidie and multiplane support is always increase performance, reduce the data access average time, and increase parallel execution of commands

[7]. NAND flash devices are programmed on a page by page basis. Typically programming time is a fe w hundred micro second per page. NAND flash cell can be programmed and erased only for limited time period (100,00 0 times for SLC and 10,000 times for MLC) before it fails. To improve this limitation, flash memory performance has b een increased by using wear leveling technique. The technique spreads the memory cell use evenly to different physi cal pages. So the entire flash device is used equally to im prove the life of flash memory [2]. Bad block manageme nt technique and wear leveling technique use some remapping technique of logical to physical address of the memory device. They all provide an FPGA to facilitate the implementation of a wide range of NAND Flash Memory Controller [4].

III. SIMULATION AND RESULT

A. Single Micro cell Module

In the simulation result, the NAND flash memory cell is simulated using Xilinx ISE Software and modelsim simulator. As shows in Figure 2, the RTL view of NAND Flash memory cell has been gene rated after the synthesis. The modelsim waveform is displayed on Figure 3 indicating the write and read operation. Write operation takes place only when both word line and write enable are '1'. The q and qbar line are modified when this situation occurs as seen in the waveform and Read operation takes place when both word line and read enable sh uld be high so that read out shows the data stored in q.

Figure 2. RTL Schematic View of Single Micro Cell

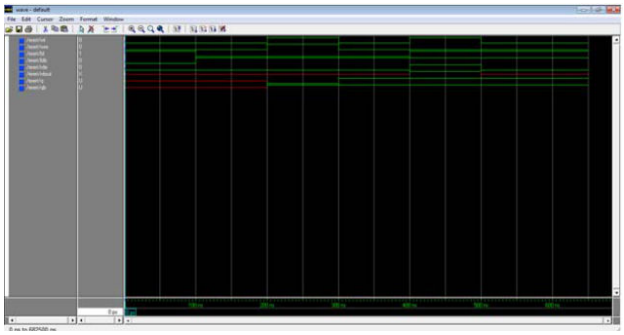


Figure 3. Simulation Waveform of Single Micro Cell

B. Full Adder

The simulation of full adder carried using Xilinx ISE Software and modelsim is shown in Figure 4. The RTL view of full adder was generated after the synthesis and it displays the internal architecture of full adder. The modelsim waveforms are displaye d in Figure 5 and depict the sum and carry output. Sum tak es place only when any one input is high and carry operation is done when at least two inputs are high.

Figure 4. RTL Schematic View of Full Adder

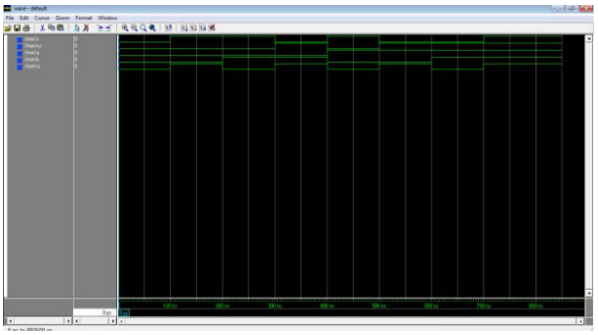


Figure 5. Simulation Waveform of Ful l Adder

C. Arithmetic and Logical Unit (ALU)

In the simulation, results of Arithmetic and Logical Unit (ALU) have been used in Xilinx ISE Software and modelsim simulator as shown in Figure 6. The RTL view of ALU has been generated after the synthesis and it displays the internal architecture of ALU. The mod elsim waveform is displayed in Figure 7 which indicates all mathematical operations. The addition operation takes place only when op_sel. Operation select lines are zero a nd provide input value to the dat_a and dat_b and chec k the results on dat_out as well as all flag register like carry flag, Sign flag, auxiliary carry flag, parity flag, etc. Thus all mathematical operation can be carried by changing the value of op_sel.

Figure 6. RTL Schematic of AL U

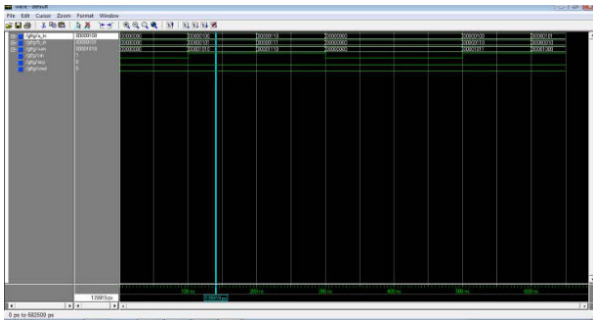


Figure 7. Simulation Waveform of ALU

D. Memory Module

In the simulation, result of Me mory Unit using Xilinx ISE Software and modelsim simulator are shown in Figure 8. The RTL view of memory unit is generated after the synthesis to display the internal architecture of Memory Unit. The modelsim waveform is displayed in Figure 9. It indicates how the data can store in to the memory with the help of read and write operation.

Figure 8. RTL Schematic of Memory Module

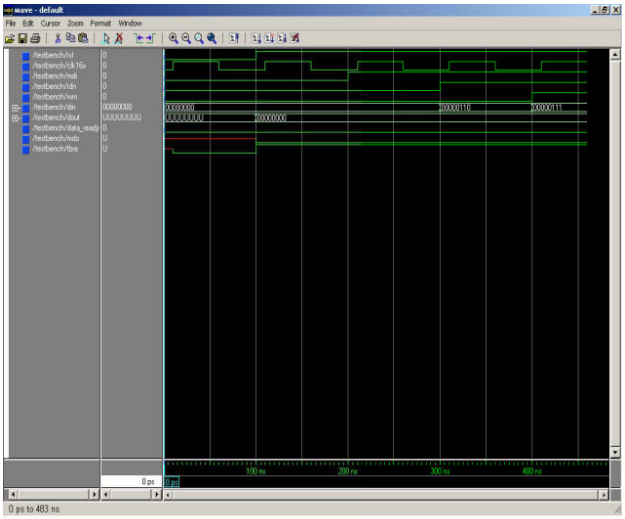


Figure 9. Simulation Waveform of Memory Module

E. Memory Read Cycle

In Figure 10, the RTL view of Memory Read Cycle has been generated after the synthesis to display the internal architecture of Memory Read Cycle. The modelsim waveform is displayed in Figure 1 1. During Memory Read Cycle when reset = '1' then both the mode operation are zero and when apply the clock pulse the data at ¼ and ¾ points in data cell is generated and when increment the clock the data are in a position to g o from serial to parallel conversion. After some time it will also shift to data register. In this way, data are shifted to the particular data register and finally read the data at that location.

Figure 10. RTL View of Memory Read Cycle

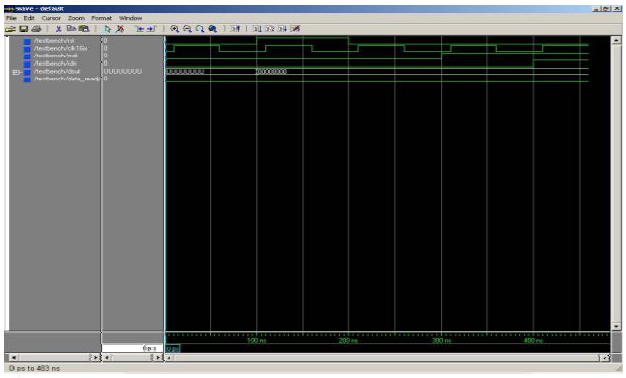


Figure 11. Simulation Waveform of Memory Read Cycle

F. Memory Write Cycle

In Figure 12, the RTL view of Memory Write Cycle has been generated after the synthesis to display the internal architecture of Memory Write Cycle. The modelsim waveform is displayed in Figure 13. During Memory Write Cycle when reset = '1' then both the wrn1 and wrn2 operation is one and when enable the clock detect the edge on write pulse. After some time, detect edge on write pulse to load transmit buffer. When increment the clock the data are in a position to go from transmit shift register to transmit buffer. In this way, data are shifted to the particular data register and finally write the data at that location.

Figure 12. RTL Schematic of Memory Write Cycle

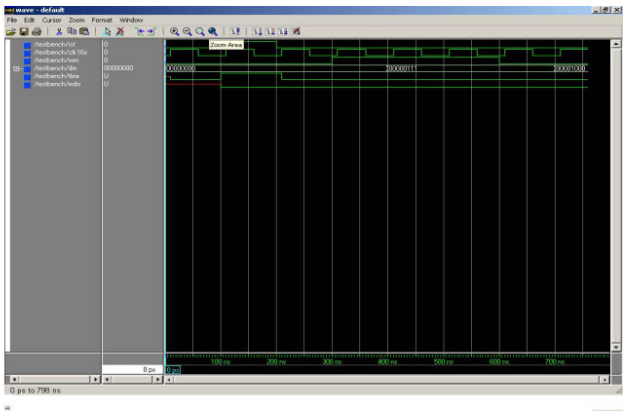


Figure 13. Simulation Waveform of Memory Write Cycle

G. NAND Flash Memory Controller

Figure 14 shows Xilinx RTL Schematic view of NAND Flash Memory Controller. Figure 15 are shows the Simulation Waveform of NAND Flash Memory Controller. We bind all the above components inside this module. This module is stand at the top of hierarchy. Various pins are described below-:

- nand_ale : This pin indicates NAND flash advanced latch enable when this pin is high, the NAND flash memory controller work and when it is low, the NAND flash memory controller are in latch mode.
- nand_ce : This pin indicates NAND flash chip enable. When this pin is high the NAND flash memory controller are in working mode, it reflects that the data is storing in a memory in sequential order. When it is low, the NAND Flash memory controller do not worked.
- nand_cle: This pin indicates NAND flash clear mode. When this pin is high the data are cleared in the main memory of NAND flash memory controller. When it is low it will not worked.
- nand_re: This pin indicates NAND flash reset. When this pin is high the NAND flash memory controller are in reset mode i.e. all operations of NAND flash memory controller stopped. When it is low, it will work properly.
- nand_tri_en: This pin indicates NAND flash tri-state enable. When it is high, it will indicate the data are in tri-state i.e. high impedance state. When it is low it may work properly.
- nand_we : This pin indicates NAND flash in write operation. When it is high, data are in writing process in a write mode i.e. we can write the data at particular memory address. When it is low, it will not provide a permission to write the data.
- nand_opr<2:0> : This pin indicates selection between reset state (000), read ID state (001),read

page state (010) and write page state (011). After that state 100,101,110,111 (4 states) reserve for future extension.

- nand_rdy : This is the NAND flash ready input signal from the memory device, only accessible when high. If it is low it will not be accessible.
- prog_2,prog_3,…prog_24 :These 23 pins used for write states
- read_2,read_3,…read_21 : These 20 pins used for read states.
- rid_2, rid_3, …. , rid_9 : These 8 pins used for select ID states.
- rst_low:This pin indicates the reset input of the system. It is normally low.
- rst_1, rst_2, rst_3 : These 3 pins used for reset states.

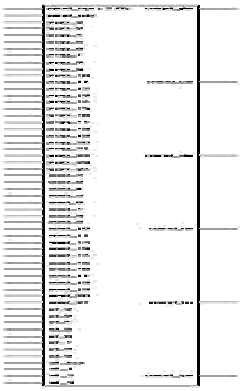


Figure 14. RTL Schematic of NAND Flash Memory Controller

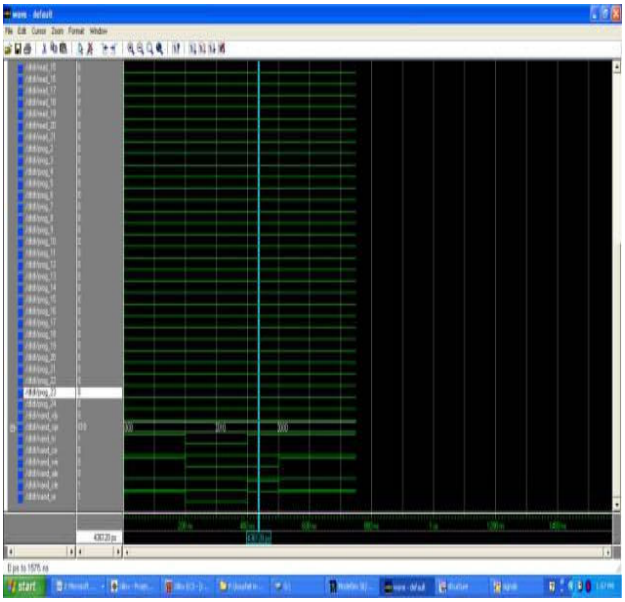


Figure 15. Simulation Waveform of NAND Flash Memory Controller

CONCLUSION

In this paper NAND flash memory controller for SD/MMC memory card using FPGA was designed. The test results show that NAND flash memory controller architecture will achieve a high performance. In the proposed NAND flash memory controller, all the blocks like microcell, microcontroller, ALU, full adder, memory unit, memory read cycle, memory write cycle have been developed. Proposed architecture can work with all the embedded computing system as a replacement of conventional Hard Disk Drive (HDD) with a very huge size of NAND flash memory. In this controller we design only four states reset state, read ID state, read page state, write page state. In future extension we have four reserve states.

ACKNOWLEDGMENT

This work was supported by ShriRam Collage of Engineering and Management, Banmore.

REFERENCES

[1] Shawn Kung, “Native PCIs SSD Controllers Next Generation Enterprise Architecture for Scalable I/O Performance”, White Paper Marvell Jan-2012.

[2] Yang Ou, Nong Xiao, “A Scalable Multi-channel Parallel NAND Flash Memory Controller Architecture”, 6th Annual IEEE Chinagrid Conference, Page 48-53, 2011.

[3] Yu Cai, Haratsch, “FPGA Based Solid State Drive Prototyping Platform”, 19th IEEE Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM), Page 101-104, 2011.

[4] Yu Chi, Erich F. Haratsch, Mark McCatney and Ken Mai “FPGA-Based Solid State Drive Prototyping Platform,” In IEEE International Symposium on Field Programmable Custom Computing Machines 2011.

[5] Yang Ou, Nong Xiao and Mingche Lai “A Scalable Multi-channel Parallel NAND Flash Memory Controller Architecture,” Sixth Annual ChinaGrid Conference 2011.

[6] International Technology Roadmap for Semiconductor, 2007.

[7] Micron Technology, Boise, Idaho. Technical Note TN-29-17 “Design and Use Considerations for NAND Flash Memory,” 2006.

[8] Altera Corporation, Stratix data isheet, May 2003.

[9] Intel Corp., Understanding the Flash Transfer Layer (FTL) Specification, 1998.

[10] Hoeseung Jung, Sanghyuk Jung, “Architecture Exploration of Flash Memory Storage Controller through a Cycle Accurate Profiling”, IEEE Transactions on Customer Electronics, Volume: 57, Page 1756–1764, 2011.

[11] Teoh, J.N., Wong W.E., “FPGA Implementation of Nonlinear Control on Hard Disk Drive”, IEEE International Conference on Control and Automation, Page 609- 614, 2009.

ADVANCED VIDEO CODING/H.264 ENCODING AND DECODING PROCESS WITH IPPPP SEQUENCES

1Anitha Perla 2Chella Moses
1*Assistant professor, Dept. of ECE, MRCE, Maisammaguda, Hyderabad
Assistant professor, Dept. of ECE, MRCE, Maisammaguda, Hyderabad
Anithakrishna77p@gmail.com,moses.chellaece@gmail.com
India

***Abstract*—** As the world expanded around us and increased the popularity of the Internet by sending receiving, uploading or downloading the high definition videos, it was necessary to use a good technology to reduce the size of the dedicated video and specialized high-quality one. If the videos are send or receive, they need a wide bandwidth to capture this amount of information in the video. Based on the above, the H.264/AVC is a good video coding standard that gives great results for encoding and decoding videos in terms frames. This technology was developed jointly by (ITU-T) International Telecommunication Union–Telecommunication Standardization, and (ISO) International Organization for Standardization. Our work involves applying the encoding and decoding process of the standard using MATLAB TOOL. The work is focusing in inter frame prediction using the (IPPPPP) frame pattern. The video that was subjected to encoding and decoding processing is using with different video coding standards such as H.264/AVC.

Keywords—IPPPP, High Quality, video, encoding, decoding

I. INTRODUCTION

H.264/AVC is the international video coding standard. It is coding. approved by ITU-T as Recommendation H.264 [1] and by ISO/IEC as International Standard MPEG (Motion Picture Expert Group)-4 part 10 AVC[2]. It is widely used for the transmission of Standard Definition (SD) and High Definition (HD) TV signals over Satellite, cable, and terrestrial emission and the storage of high-quality SD video signals onto DVDs[3]. However, an increasing number of services and growing popularity of high definition TV are creating [4] greater needs for higher coding efficiency. Moreover, other transmission media such as Cable Modem[5] offer much lower data rates than broadcast channels, and enhanced coding efficiency can enable the transmission [6] of more video channels or higher quality video representations within existing digital transmission capacities[7]. Digital Video is a sequence of still images or frames and represents scenes in motion[8]. A video signal is a sequence of Two Dimensional (2D) images projected from a dynamic three dimensional (3D) scene onto the image plane of a video camera[9]. Video coding is the process of compressing and decompressing a digital video signal[10]. Coding of video is performed picture by picture. Each picture to be coded is first partitioned into a number of slices[11]. Slices are individual coding units in this standard as compared to earlier standards as each slice is coded independently[12]. The hierarchy of video data organization is as follows: picture-slices-macroblocks-sub macro blocks-blocks-pixels[13]. The main elements of the H.264, source coder are prediction, block transformation (spatial to frequency domain translation), quantization, and entropy[14-16].

II. LITERATURE SURVEY

The purpose of this section is to provide a short description of the various video compression standards. Standards allow interoperability between different manufacturers and a variety of equipment worldwide. Two standards bodies have strongly influenced the development of video compression standards: the International Telecommunications Union (ITU) and the Moving Pictures Experts Group (MPEG)[3-5]. The oldest body is the ITU .Of particular interest is the Telecommunication Standardization Sector, or ITU-T sector[11]. Their goals have been transmitting video over both the analog and digital telephone system although in recent years the difference between their goals and ISO MPEG are very much blurred[16]. Video compression standards developed by ITU are designated by the label"H.26*X*"[17].

In the early 1990's the International Organization for Standardization (ISO) began looking at video compression for computer and multimedia applications. ISO formed MPEG to develop video compression standards for ISO. MPEG can be thought of as the "computer guys." Standards developed by MPEG are designated by the label "MPEG-*x*." The MPEG committee tends to be better known than the ISO although both had an equally important impact on the video compression industry[18]. It is important to note that both ITU and MPEG video compression standards only describe the decoder and not the encoder structure[19]. The standards describe the syntax of the encoded bit stream as well as behavior of a compliant decoder. Developers then are allowed to design the encoder anyway they want as long as it produces a compliant bit stream[21-21].

In order to achieve better compression, the video may be pre-processed before being processed by the encoder. Since errors may be introduced to the compressed video data, error conceal mentor block-artifact reduction techniques may be used after the decoding process to enhance the overall quality of the video. It should be emphasized that such techniques are not part of the video compression standards.

A timeline of the development of video compression standards is shown in Figure 2.2, with the

contributions of MPEG and ISO shown above and below respectively. It should be noted that MPEG and ITU jointly developed two standards, these are MPEG-2/H.262 and MPEG-4: Part10/H.264.

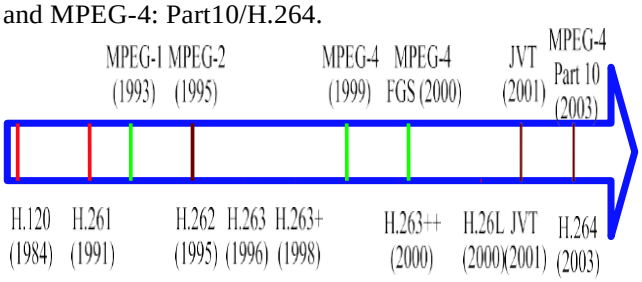


Fig. 2.2. Timeline of video coding standards. MPEG-2/H.262 and MPEG-4: Part 10/H.264 were joint projects of MPEG and ISO.

H.120 Standard:

H.120 was the first digital video compression standard. It was developed by COST 211. The video turned out not to be of adequate quality, there were few implementations, and there are no existing codecs for the format, but it provided important knowledge leading directly to its practical successors, such as H.261.The drawbacks are H.120 video was not of good enough quality for practical use it had very good spatial resolution (as differential PCM works on a pixel-by-pixel basis), but very poor temporal quality. It became clear to researchers that to improve the video quality without exceeding the target bitrate for the stream, it would be necessary to encode using an average of less than one bit for each pixel.

H.261 Standard:

H.261 is an ITU-T video compression standard, first ratified in November 1988. It is the first member of the H.26x family of video coding standards in the domain of the ITU-T Video Coding Experts Group (VCEG), and was the first video coding standard that was useful in practical term. H.261 was originally designed for transmission over ISDN lines on which data rates are multiples of 64 kbit/s. There are however a few difficult problems in H.261: Motion vector search, Bit-rate Control, Propagation of Errors.

H.262 Standard:

H.262or MPEG-2 Part 2 (formally known as ITU-T T Recommendation H.262 and ISO/IEC, also known as MPEG-2 Video) is a video coding format developed and maintained jointly by ITU-T Video Coding Experts Group (VCEG) and ISO/IEC Moving Picture Experts Group (MPEG). MPEG-2 Video is similar to MPEG-1, but also provides support for interlaced video (an encoding technique used in analog NTSC, PAL and SECAM television systems). MPEG-2 video is not optimized for low bitrates (less than 1 Mbit/s), but outperforms MPEG-1 at 3 Mbit/s and above. All standards-conforming MPEG-2 Video decoders are fully capable of playing back MPEG-1 Video streams.

H.263 Standard:

H.263 is a video compression algorithm and protocol which is standardized by ITU. It is due to be published sometime in 1995/1996. It was designed for low bitrate communication, early drafts specified data rates less than 64 Kbits/s, and however this limitation has now been removed. It is expected that the standard will be used for a wide range of bitrates, not just low bitrate applications, and expected that H.263 will replace H.261 in many applications. The Video source coding algorithm of H.263 is based on Recommendation H.261 and is a hybrid of inter-picture prediction to utilize temporal redundancy and transform coding of the remaining signal to reduce spatial redundancy, however with some changes to improve performance and error recovery. So, In comparison with video compression H.261 which is widely used for ISDN video conferencing, H.263 can achieve the same quality as H.261 with 30-50% of the bit usage. Most of this is due to the half pixel prediction and negotiable options in H.263. H.263, in addition, is also better than MPEG-1/MPEG-2 for low resolutions and low bitrates.

H.264 / MPEG-4 Part 10: Advanced Video Coding

In early 2000 ITU and MPEG began working jointly again on a new video com- pression standard. They formed a group known as the Joint Video Team (JVT) to examine new issues in video compression . The offcial ITU and MPEGdesigna- tions are H.264 and MPEG-4 Part 10: Advanced Video Coding respectively. In this thesis, the ITU designation H.264 will be used to refer to the standard.

This standard was designed to target a wide variety of applications, including wireless , IP networks , and digital cinema.The standard has defined two separate main coding layers: the Video Coding Layer (VCL) and the Network Abstraction Layer (NAL). The VCL will be the focus of this paper, although the NAL will be discussed in some detail.

In comparison to previous standards, this standard is a departure from earlier standards, introducing many new technical features to further increase compression effciency, such as flexible macroblock sizing, 1/4-pixel interpolation , multiple reference picture capabilities, and an in-loop filter.

III.PROPOSED WORK

Encoder Decoder:

The objective of image processing is to analyze image data to make the system to understand, recognize and interpret the processed information available from the image pattern. For example, noise removal, sharpening, or enhancing brighten an image, making it easier to identify key features. These enhancement can be classified into two main groups – spatial domain based and transformation domain based methods . The popular histogram equalization (HE) based methods come in to the spatial domain, which can be further classified into two categories, i.e., global and local histogram equalization based techniques. Global Histogram Equalization (GHE) methods explore the histogram information of the entire image to form its transformation function. The classical HE can efficiently utilize display intensities, but it tends to over enhance the contrast if there are high peaks in the Histogram, which

often results in a harsh and noisy appearance of the output image.

The video can be divided in to number of frames, that sequence can me in the form of f1,f2,f3…….fn-1,fn.Then again frame can be divided in to macro block is encoded in an inter and intra mode can be perform with motion estimation and motion compensation that is to be count as prediction process.

In this work inter prediction mode is performing with motion compensation prediction with previous frame that is to be consider as reference frame. That the sequence to be consider as IPPPPPPP----.

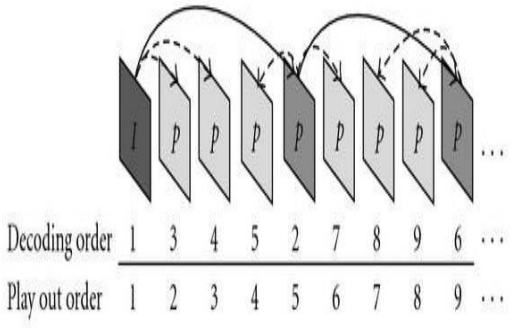


Figure. 2. IPPPP--- sequence

The prediction picture is subtracted from the current block to produce as residual block that process can be consider as transformation process. Here transformation technique has implemented with hybrid wavelet transform. And quantized to be a set of coefficients i.e quantized transform coefficients, which are to be ordered and apply entropy process.

H.264 decoder is inverse process of H.264 encoder and it is used to decode the encoded video come from the encoder. Decoder receives compressed bit stream and decodes data elements to produce set of quantized coefficients, these coefficients has to be decoded from the bit stream and decoder creates prediction block, from that prediction block it has to send to Deblocking filter and generate frames.

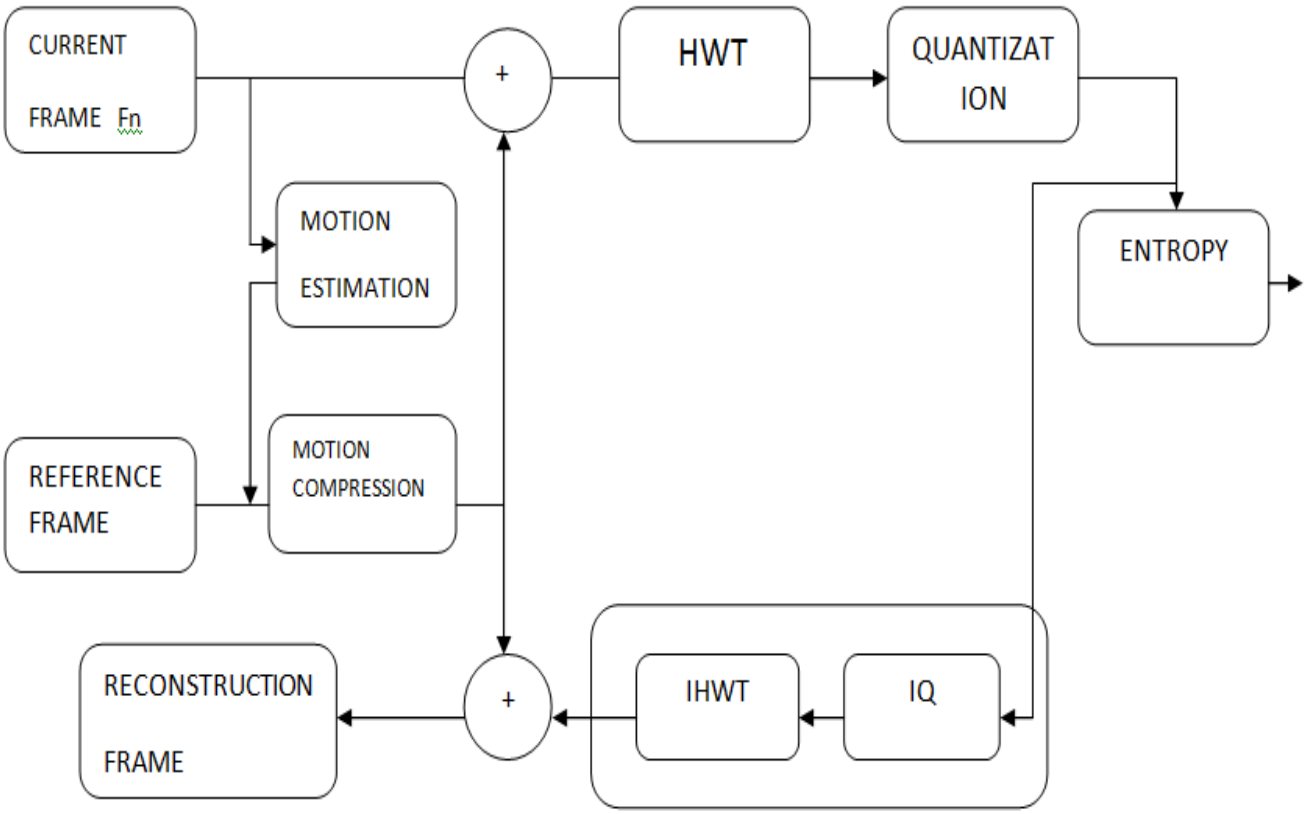


Figure 3 Proposed Block Diagram Of H.264

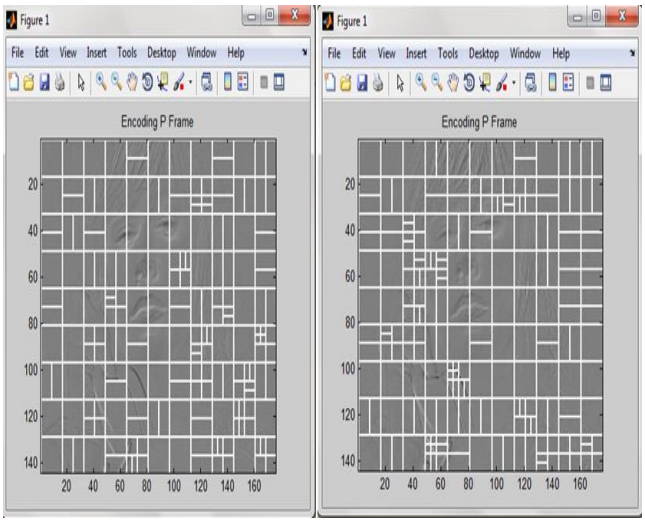
In figure 3 represents the encoder and decoder of the proposed techniques. It gives the detail explanation of the process of the video sequence uses the intra and inter frame formats in terms of compression efficiency at the time of writing (early 2010). The Moving Picture Experts Group (MPEG) and Video Coding Experts Group (VCEG) are examining the need for a new video compression standard. The consensus was that (a) there is likely to be a need for a new compression format, as consumers demand higher-quality video and as

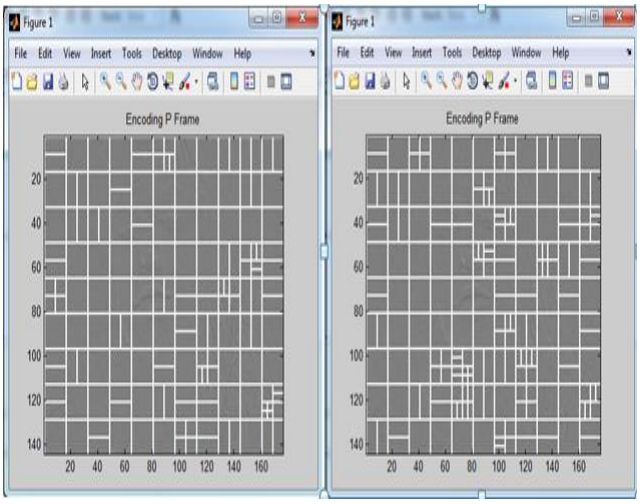
processing capacity improves and (b) there is potential to deliver better performance than the current state-of-the art. A number of different techniques were proposed, including decoder-side motion estimation, larger macro block sizes (up to 32 × 32), and more sophisticated in-loop Deblocking filters, adaptive transform sizes and improved intra prediction. In general, all of these proposed algorithms offer the potential for better compression performance at the expense of increased computational complexity. Results of subjective comparison tests of the proposed architecture to conclude that 'for a considerable number of test sequences significant peak signal noise ratio (PSNR) could be achieved'. prediction, later on motion Compensation and motion Estimation is done on the frames by using the reference frame.

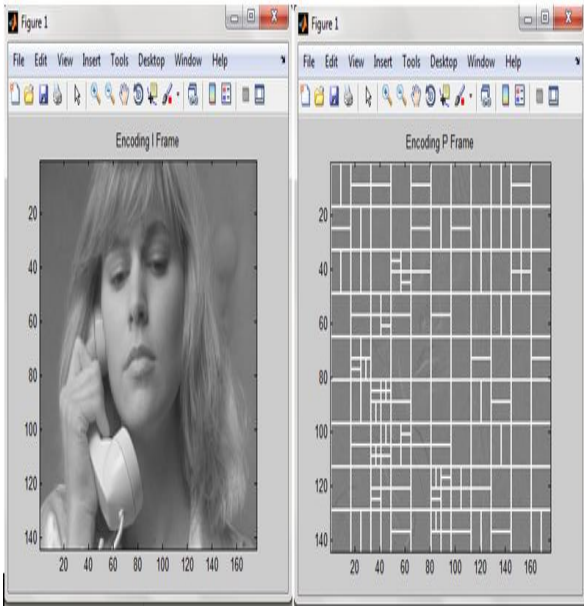
IV. SIMULATION RESULTS

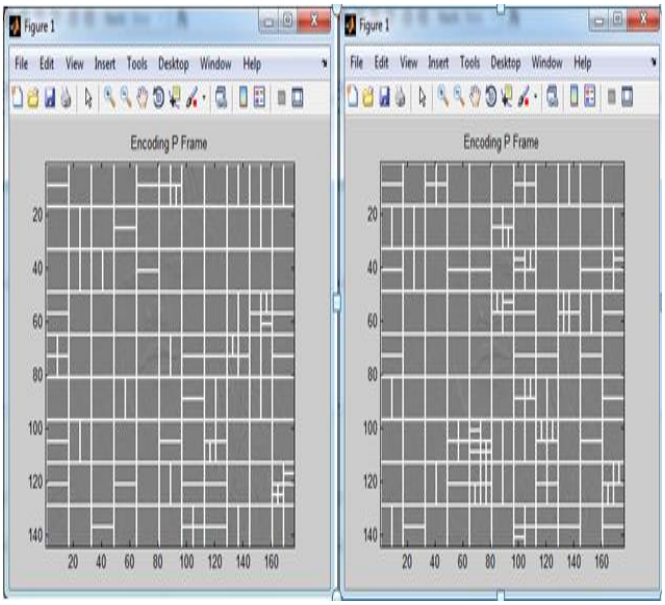
Encoding Process:

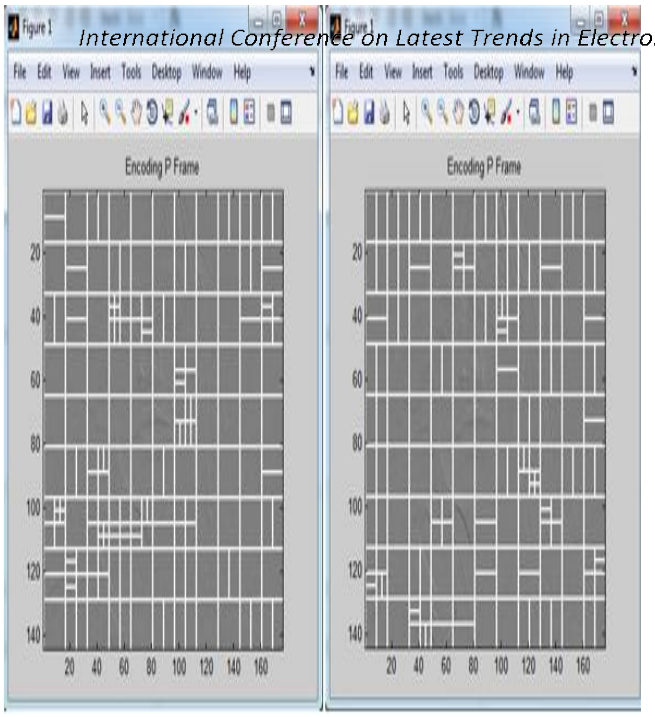
In this process 'suzie_qcif.yuv' is considering as a input video sequence. The block diagram of the basic encoder with extensions for non-intra frame coding techniques is given. Of course, this encoder can also support intra frame coding as a subset. Starting with an intra, or I frame, the encoder can forward predict a future frame. This is commonly referred to as a P frame, and it may also be predicted from other P frames, although only in a forward time manner. As an example, consider a group of pictures that lasts for 10 frames. In this case, the frame ordering is given as I, P, P, P, P, P,I,P,P,P,P,....sequence.

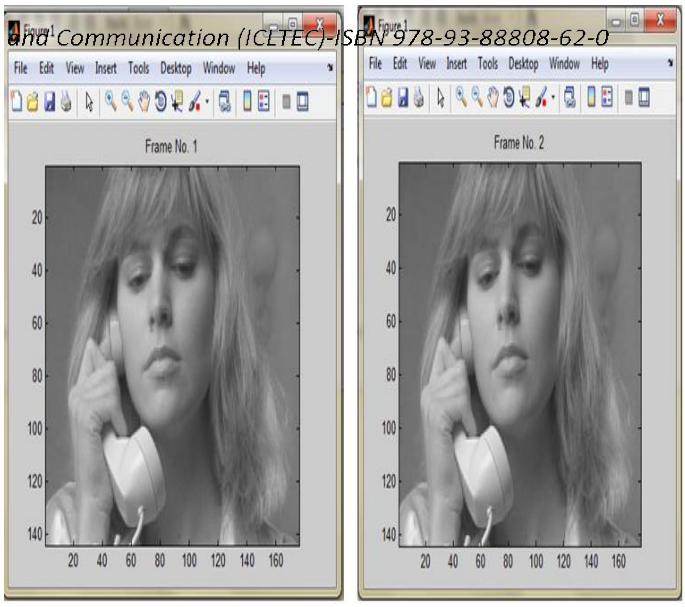


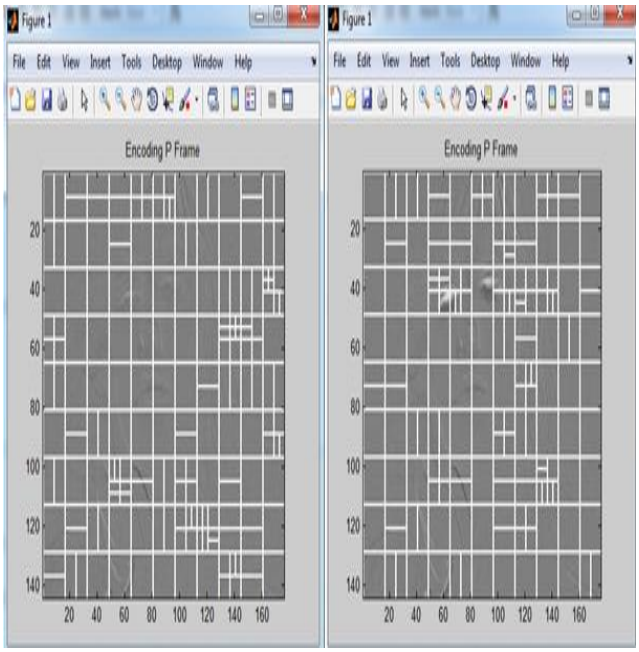


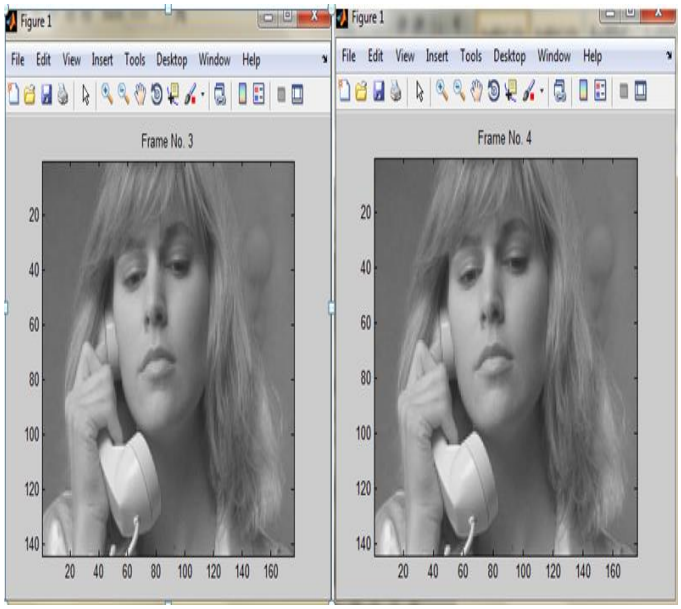


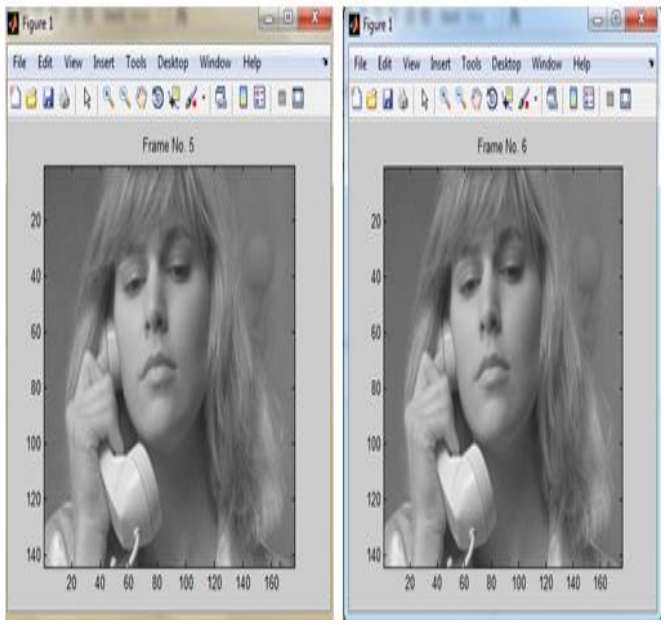






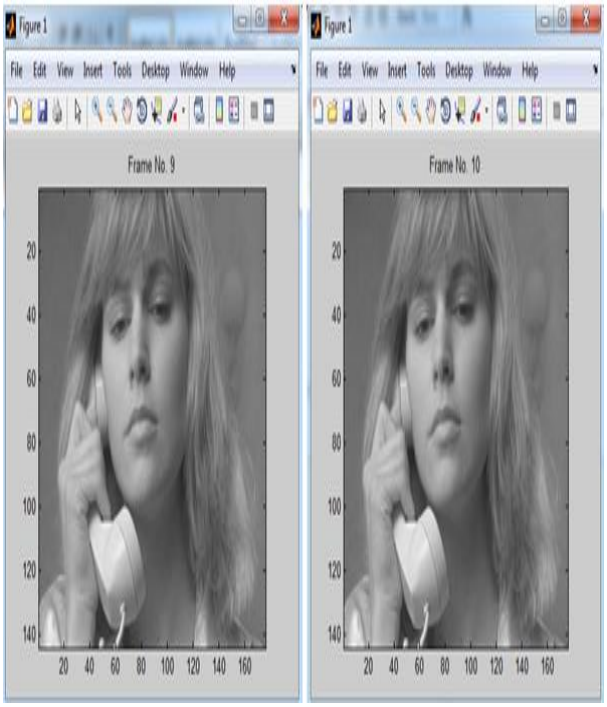






Decoding process:

In this process bit stream, The block diagram of the basic decoder with extensions for non-intra frame coding techniques is given. Of course, this decoder can also support intra frame coding as a subset. Starting with an intra, or I frame, the encoder can forward predict a future frame. This is commonly referred to as a P frame, and it may also be predicted from other P frames, although only in a forward time manner. As an example, consider a group of pictures that lasts for 10 frames. In this case, the frame ordering is given as Frame 1 , Frame 2 ,... Frame 10.



First standardized in 2003, H.264/AVC is now a relatively mature technology. H.264 is certainly one of the leading subjective quality of one point, his implies that there is scope for a new coding format that significantly out-performs H.264/AVC. The current plan is to set up a Joint Collaborative Team (JCT) of MPEG and VCEG representatives to work on a new video coding standard. In this work, considered 'suzie_qcif.yuv' as test sequence its size is 176 width as well as 144 height, frame frequency is 30 Hz (30 frames per sec). The GOP test sequence is IPPPPPP-----, in this process I frame is having complete information of the scene or content of the image, whereas following P sequences is completely using comparison with I frame with process of motion estimation and compensation accounted as motion prediction technique such as inter and intra prediction. The split procedure partitions the MB into variable size block using a quad tree approach. In this method a macro block is divided into quarters of equal area Then using similarities of motion vectors of adjacent blocks we will show how to merge the sub-blocks for quarter division Loading a file and then generate the bit stream, the regular encoding process is to be done with motion estimation with different quantization parameters (QP) value. The same in the reverse process can be done in the Decoder, encoding frames and decoding frames As shown in the below

CONCLUSION

H.264/AVC represents a major steps in the development of video coding standards, in terms of both coding efficiency enhancement and flexibility for effective use over a broad variety of network types and application domains. In this work survey has been done in terms all the technical features such as Transformation in terms of various methods, motion estimation and compensation in terms of Inter and Intra Prediction and also the final section is Deblocking filtering process. Among them is enhanced motion prediction capability, use of small block-size exact-match transform, adaptive in-loop de-blocking filter, and enhanced entropy coding methods. The H.264/AVC is highly flexible with all motion model and the very efficient performance as comparing with the existing methods (H.261/H.262/H.263) performance of H.264/AVC as described in the project.

REFERENCES

[1]. T. Wiegand, G. J. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Sys-temsforVideoTechnology*,vol.13,no.7,pp.560–576,July2003.
Samitha T,Study of H.264/Avc Algorithm and It's Implementation In Matlab,jan 2014.
[2]. Thomas Stutz, A Survey of H.264 AVC/SVC Encryption. VOL. 22, NO. 3, MARCH 2012.
[3]. Mohammed A. Saleh, Coding Structure and Performance of High Efficiency Video Coding (HEVC) and H.264/AVC.
[4].I.E.Richardson,*H.264andMPEG-4VideoCompression*.Hoboken,NJ:Wiley, 2003.
[5].K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms,Advantages, Applications*. Boston, USA: Academic Press,1990.
[6]. Damián Ruíz, fast algorithm for h.264/avc intra prediction based on discrete wavelet transform.
[7]T.Elarabi, hybrid wavelet - dct intra prediction for h.264/avc interactive encoder.
[8]Pengyu Liu, A fast and novel intra and inter modes decision prediction algorithm for H.264/AVC based-on the characteristics of macro-block,2009
[9]Ashika K O, Inter Frame Coding in Advanced Video Coding Standard H.264 using Block Based Motion Compensation Technique,2017.
[10]M. Chan, Y. Yu, and A. Constantinides, "Variable size block matching motion compensation with applications to video coding," *IEEE Proceedings on Com- munication, Speech, and Vision*, vol. 2, no. 4, pp. 205–212, Aug.1990.
[11]Asim Munir, Deblocking Filters for Block-Based ICA Transform Coding,2010.
[12].Nikolaos Kefalas, a parallel luma-chroma filtering Architecture for h.264/avc deblocking filter,2015.

[13]Y. Lee and H. Park, "Loop filtering and post-filtering for low-bit-rates moving picturecoding,"*SignalProcessing:ImageCommunications*,vol .16,no.9,pp. 871–890, June2001.
[14]Tsu-Ming Liu, An In/Post-Loop Deblocking Filter With Hybrid Filtering Schedule,2007.
[15] Vinayak Ashok Bharadi, Hybrid Wavelet Transform I and II Combined with Contrast Limited Adaptive Histogram Equalization for Image Enhancement,2015.
[16] M. Krupa Swaroopa Rani, Implementation Of Image Compression Using Hybrid Dwt -Dct Algorithms.
[17]U. S. Mohammed and W. M. Abd-elhafiez, "Image coding scheme based on object extraction and hybrid transformation technique," Int. Journal of Engineering Science and Technology, vol. 2, no. 5, pp. 1375–1383, 2010.
[18]Suchitra Shrestha and Khan Wahid, "Hybrid DWT-DCT Algorithm for Biomedical Image and Video Compression Applications",2010.
[19].A. Alice Blessie1, J. Nalini and S.C.Ramesh, "Image Compression Using Wavelet Transform Based on the Lifting Scheme and its Implementation", May2011.
[20]Subhasis Saha, "Image compression - from DCT to Wavelets- AReview",2009.
[21]A. Puri, H. Chen, A. Luthra, "Video Coding using the H.264/MPEG-4AVC compression standard," *Signal Processing: Image Communication*, October 2004.
[22]Yu-Wen Huang, Analysis and Complexity Reduction of Multiple Reference Frames Motion Estimation in H.264/AVC,Apiril 2006.
[23]Jamil-ur-Rehman,Efficient Techniques for Signalling Intra Prediction Modes of H.264/Mpeg-4 Part 10,2006.
[24]Digital Image Processing, Rafael C. Gonzalez ,2009.

Low Power Viterbi Decoder Design based on Reversible Logic Gates

Nenavath Naresh

PG Student, Dept. of ECE, MRCE, Hyderabad, India, Email: nareshnenavath4@gmail.com

***Abstract*— In recent trends of VLSI technology the reversible logic has became the major area of research in optimization of area, power and speed constraints. The reversible logic has equal number of inputs and outputs. In wireless communications Viterbi algorithm is employed to have minimal number of communication channels. The Viterbi decoder design at 65nm technology using reversible logic has made an attempt for optimizing power, area and delay with increased efficiency.**

***Keywords*—** *Viterbi decoder, Reversible logic, Feynman Gate, Peres Gate, HN Gate, Power consumption, Delay.*

I. INTRODUCTION

The Viterbi decoding algorithm was proposed by Andrew J Viterbi, which is a decoding process for convolutional codes in memory-less noise. This algorithm is implemented in the designing of communication systems. The Viterbi Algorithm is the most resource consuming and it finds the most-likely noiseless state transition sequence in a state diagram, given a sequence of symbols which are interrupted by noise [1].

Generally, a viterbi decoder consists of three basic computation units: Branch Metric Unit (BMU), Add-Compare-Select Unit (ACSU) and Survivor Memory Unit (SMU) [2].

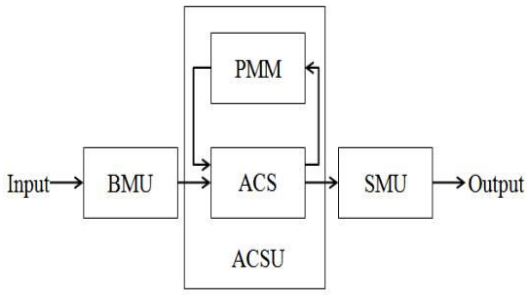


Figure 1. Block diagram of Viterbi decoder

The primary unit is called as Branch Metric Unit (BMU). This unit will compare the received data symbols with the ideal outputs of the encoder and finally the branch metric will be calculated. The Euclidean distance or Hamming distance is utilized for the calculation of branch metric. The BMU creates branch metrics for the following module in terms of the symbols which are received by the channel.

The Add Compare Select Unit (ACSU) is illustrated as the sum of the Branch Metrics (BM) to the subsequent Path Metrics (PM). The new PM will be compared and the selected PM will be stored in the Path Metric Memory (PMM). At the same time, the ACSU stores the associated survivor path decisions in the Survivor Memory Unit (SMU).The PM of the survivor path of each state is updated and stored back into the PMM.

The Survivor Memory Unit uses the Trace-Back method to identify the survivor path and output data. In this unit the decoded bits are extracted from the beginning through

Minimum path metric. At the beginning state, backward tracing is followed by the survivor path, which initially contributed to the current PM and a unique path is identified. While tracing back through the trellis, the decoded output sequence corresponding to the traced branches is generated in the reverse order.

II. REVERSIBLE LOGIC

Reversible computing is the application of principles of recycling to computing. A reversible logic gate is mapped with one-to-one logic device having an n-input, n-output gate. As it helps to find out the outputs from inputs although the inputs can be exclusively recovered from the outputs. In the necessary conditions to have the number of inputs equal to the number of outputs additional inputs or outputs is added. An important constraint present on the design of a reversible logic circuit using reversible logic gate is that the fan-out is not allowed. The quantum cost of reversible logic circuits must be minimum. With the minimum number of reversible gates the design of reversible circuit is accomplished. The major constraint to attain optimization of the circuit is to produce the garbage outputs and the constant inputs with the minimum number.

The reversible logic gates are the circuits which has number of inputs is equal to number of outputs. The important optimization parameter for every reversible logic gate is the quantum cost [3].

The important reversible logic gates which are required for designing Viterbi decoder are.

A. Feynman Gate

The Feynman gate is 2x2 reversible gate which has the inputs (A, B) and the outputs (P =A, Q = A⊕ B). This gate is also referred as Controlled NOT. The quantum cost is 1. This is mainly used for the fan-out function. The power consumption and delay are 18mW and 7.760ns [4].

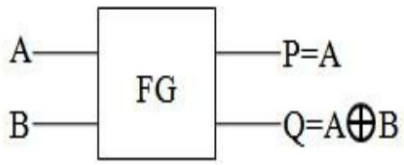


Figure 2. Feynman Gate

B. Peres Gate

The Peres gate is 3x3 reversible gate, where the inputs are A, B, C and the outputs are P, Q and R. The outputs are mapped as P =A, Q = A⊕ B and R= A.B⊕ C. The quantum cost is 4. The power consumption and delay are 24mW and 7.824ns.

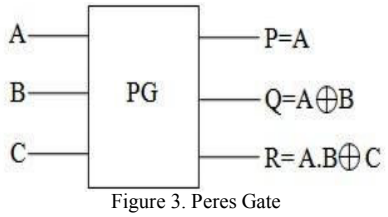


Figure 3. Peres Gate

A. HN Gate

The HNG is a 4x4 reversible gate, and abbreviated as Haghparast Navi Gate (HNG). It has 4 inputs A, B, C, D and 4 outputs P, Q, R, S, its mapping is P=A, Q= B, R= A⊕ B⊕ C and S= (A ⊕ B) C ⊕ AB ⊕ D. The quantum cost is 6. The power consumption and delay are 24mW and 7.824ns.

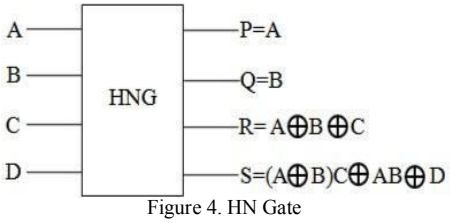


Figure 4. HN Gate

III. PROPOSED VITERBI DECODER USING REVERSIBLE LOGIC GATES

In proposed system the calculation of metric, addition, comparison of weight and selection of survivor path everything is carried out in the ACS array (ACSU). Thus the ACS array contains the values of weight at every state, as a progression along the survivor path. The codewords are very much required when comparing with different codewords of different paths, to identify the survivor path. The functions of compilation and comparison takes place within the ACS array which actually identifies the extensions of path. The signals determining these extensions to the survivor paths are passed from the ACS array to the SPU, which then updates the survivor paths.

The Figure 5 shows the top level RTL design of Viterbi decoder in which U1 indicates compute metric unit, U2 represents metric unit, U3- acs enable unit, U4- compare select unit, U5- reduce unit and U6- path memory unit.

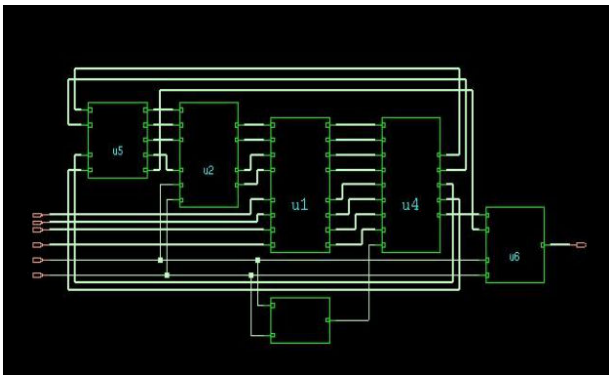


Figure 5. Top Level RTL design of Viterbi Decoder using Reversible Logic Gates

The design has the four input signals each signal having two bits with clock and reset signals. As the Reversible Logic gates reduces the power consumption and path delay, the PG will function as half adder and HNG will function as full adder.

A. Compute Metric Unit

In the compute metric unit, each received code word and each excepted code word (metric output) will be the inputs for this unit in which if performs the comparison between the code words and generates a codeword which is the input for the compare select unit.

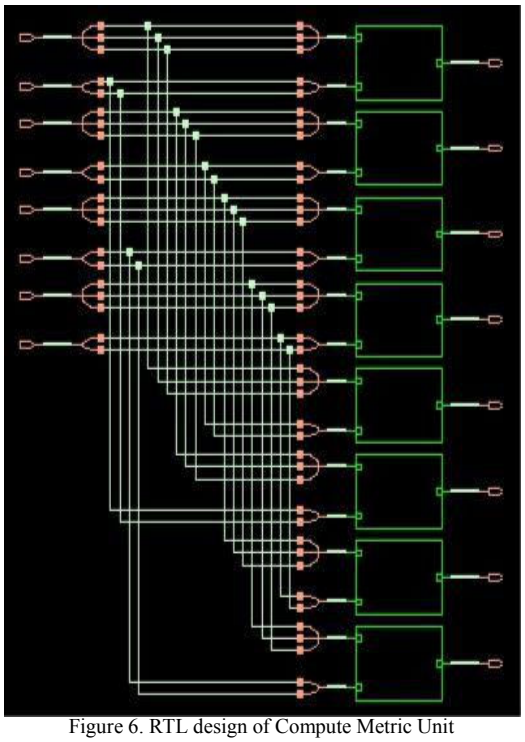


Figure 6. RTL design of Compute Metric Unit

The Figure.7 shows a unit of compute metric in which expected bits are represented as a [0:2] and received bits as b [0:1]. The blocks U_FA_0 and U_FA_2 is implemented by Peres gate which performs the half adder and U_FA_1 is implemented by HNG performing full adder.

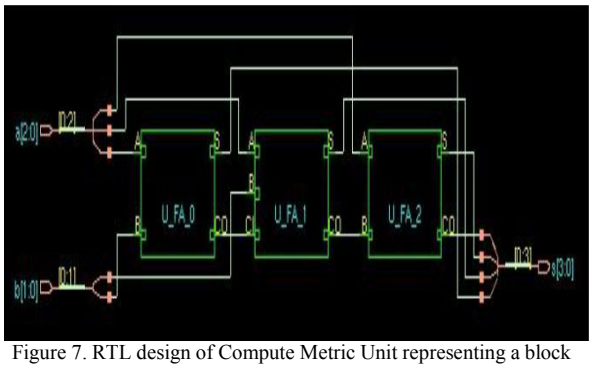


Figure 7. RTL design of Compute Metric Unit representing a block

Each bit of the metric unit will be the inputs for the 3 sub units and user input of 2 bits will be for the inputs for U_FA_0 and U_ FA_ 1. The U_FA_0 will generates the output of 2 bits, one bit will be the final output of the block and other bit will be the input for the next sub unit i.e. U_FA_1. The U_ FA_1 will generates the output of 2 bits, one bit will be the final output of the block and other bit will be the input for the next sub unit i.e. U_FA_2. Similar function is performed by the U_FA_2 sub unit and both the outputs. The unit has 8 blocks and each block performs the same function as explained above. Hence the compute metric has 8 outputs with each output of 4 bits.

B. Metric Unit

The metric unit performs the storage operation of the expected code word. The Figure 8 shows the RTL design in which it has the cock and reset signals for each metric unit.

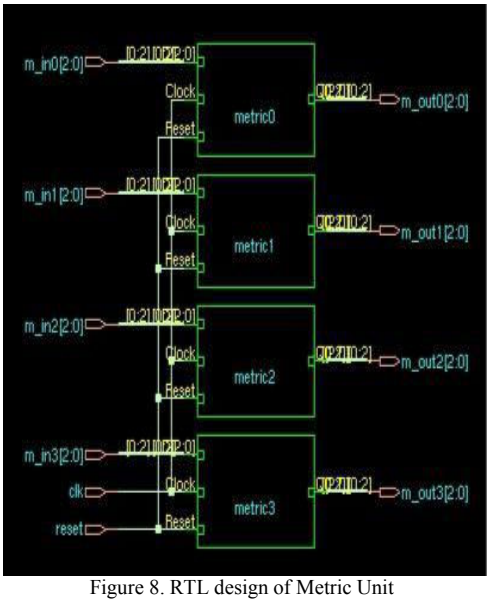


Figure 8. RTL design of Metric Unit

The metric unit stores the minimum metric of the oldest bits as it is required for next computation cycle. Every time the oldest path will be erased and the new minimum reduced metric will be updated. The unit has 4 metric blocks and each metric block has 3 inputs and 3outputs.

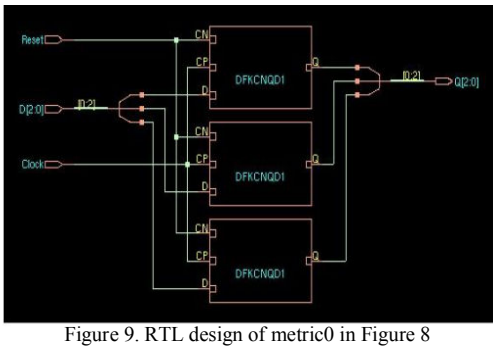


Figure 9. RTL design of metric0 in Figure 8

Figure 9 represents the each metric unit which consists of three D flip-flops, once the clock signal is applied upon reset=1 the bits will perform the comparison operation in the compute metric unit.

C. ACS Enable Unit

Figure 10 shows the acs-enable unit which consists of D flip-flops which has clock and reset signals and initially the reset will be 0, D input is 0.

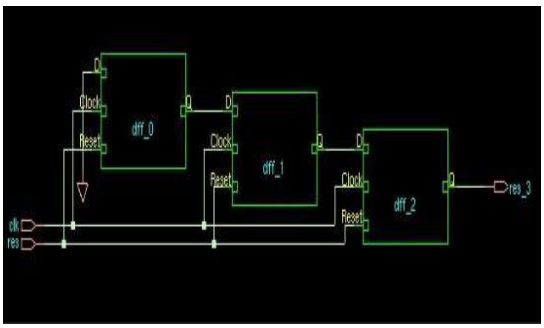


Figure 10. RTL design of ACS-Enable Unit

Once the reset value becomes 1 the unit will send a signal such that a compare-select unit is enabled.

D. Compare Select Unit

The codeword which is received from the compute metric unit will have maximum path metric such that the minimum path metric has to be calculated by selecting the appropriate metric, this operation in performed by the compare select unit and Figure 11 shows the RTL design of the unit.

Upon receiving the acs enable signal this unit will be enabled and the each received codeword from the compute metric unit will be of 4 bits. Two code words will be added and compared such that minimum metric will be calculated and obtained by performing the logical operations such as AND and OR. Finally the minimum metric will be selected and the codeword will be three bits. The output will be of four bits which will be stored in the path memory. The minimum metric which is calculated will be reduced such that previous path will be deleted. The path metric of both inputs will be compared and the metric with smallest distance will be the output and this is considered as the control signal for the path memory unit.

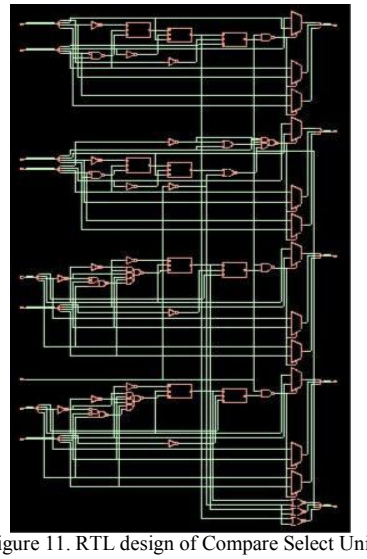


Figure 11. RTL design of Compare Select Unit

E. Reduce Unit

The minimum path metric which is obtained has to compared with the oldest path metric and it will be stored in the reduce unit, suppose if both the metric are equal then the oldest path metric will be retained otherwise the new minimum path metric will be stored in the reduce unit.

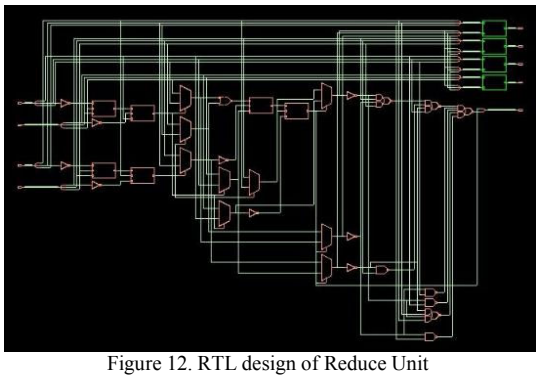


Figure 12. RTL design of Reduce Unit

The old minimum path metric which is stored will be evaluated with the HN and Feynman reversible gates such that the new minimum metric of 3 bits will be stored by eliminating the oldest path.

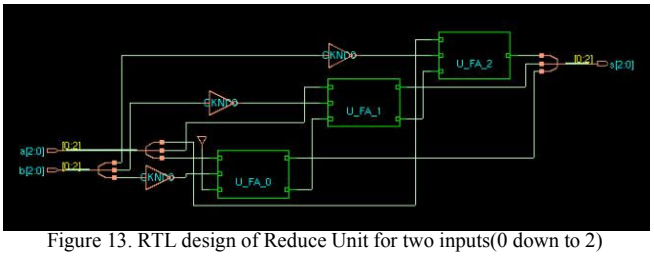


Figure 13. RTL design of Reduce Unit for two inputs(0 down to 2)

The control signal of 2 bits is required for the path memory to enable the unit and this is generated by the reduce unit by latch which consists of OR and AND logical gates.

F. Path Memory Unit

The path memory unit consists of 4x1 multiplexer units and the D flip -flops. The control signal is considered as a select signal for the multiplexer as it is used to select the minimum path metric of the decoded message bit.

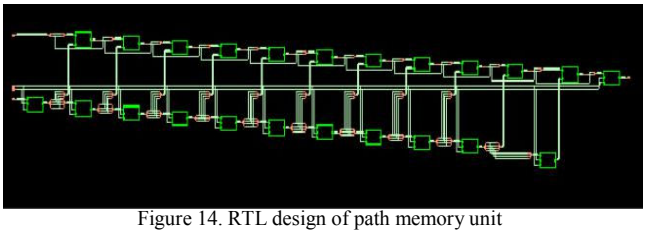


Figure 14. RTL design of path memory unit

The input for the D- flip flop is the ACS signal which is the output from the compare-select unit of 4 bits. These bits are considered as a survival bit inputs for the 4x1 multiplexer.

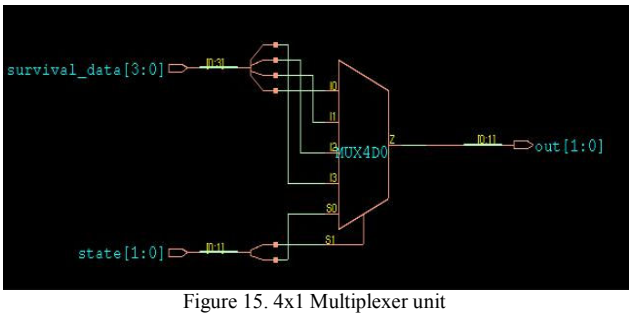


Figure 15. 4x1 Multiplexer unit

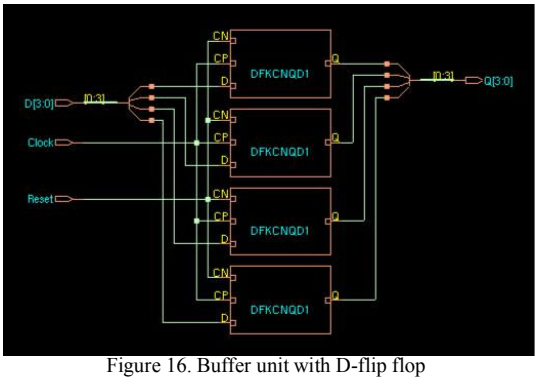


Figure 16. Buffer unit with D-flip flop

During the initial stage the path 1 is identified, then path 2 is identified by the multiplexer units. The unit consists of 11 multiplexer units and 9 D flip-flops.

The decoder output will be available at the receiver end after the completion of 14 clock cycles. Since the clock and reset signal are applied to the acs enable and path memory units.

IV. Results and Discussion

The proposed Viterbi decoder is developed by Verilog coding and simulated. The main constraints for VLSI design is area, speed and power. The power performance report is shown in Table 1, area and timing performance report is tabulated in Table 2.

Table 1. Power Performance Summary

INSTANCE	Power Analysis in ηW		
	Leakage Power	Dynamic Power	Total Power
U1 [Compute Metric]	1436.774	4871.918	6308.693
U2[Metric]	555.193	2961.352	3516.545
U3[ACS_Enable]	157.896	298.383	456.279
U4[Compute Select]	520.295	3986.532	4506.828
U5[Reduce]	1000.059	5939.455	6939.514
U6[Path Memory]	2697.978	11326.026	14024.004
VITERBI [TOP LEVEL]	**6368.196**	**30094.526**	**36462.722**

Table 2. Area and Timing Performance Summary

INSTANCE	Number of Cells	Area (sq microns)	Time Delay (psec)
U1 [Compute Metric]	40	173	320
U2[Metric]	12	86	135
U3[ACS_Enable]	3	22	0
U4[Compute Select]	51	111	533
U5[Reduce]	74	177	1343
U6[Path Memory]	57	422	2213
VITERBI [TOP LEVEL]	**237**	**990**	**4544**

The functionality of Viterbi decoder using reversible logic gates is verified and simulated. The simulation waveform result is shown in Figure 17.

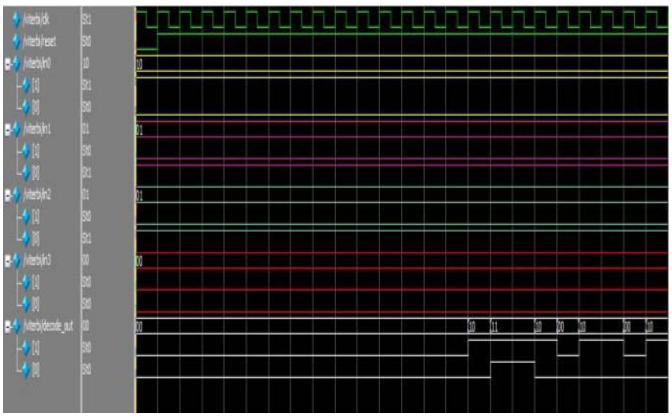


Figure 17. Schematic waveforms

Clk & reset; in0 (input); in1 (input); in2 (input); in3 (input); decode_out

V. CONCLUSION

The Viterbi decoder which consists BM, ACS and SM Units have been implemented by considering the reversible logic gates such as Feynman, Peres and HN gates to achieve a better efficiency. The designing is done using HDL coding and synthesized in CADENCE tool. The performance summary such as area, power efficiency and delay results are achieved. Hence this design can be used in high speed communication applications.

REFERENCES

[1] Abdulfattah M. Obeid, Alberto Garcia, "*A Multipath High Speed Viterbi Decoder*", IEEE 2003: pp.1160-1163.

[2] Dalia A. El-Dib and Mohamed I. Elmasry, "*Modified Register-Exchange Viterbi Decoder for Low-Power Wireless Communications*", IEEE Transactions on Circuits and Systems, Vol.51, No.2 February 2004: pp.371-378.

[3] H.R.Bhagyalakshmi, *"An Improved Design of a Multiplier Using Reversible Logic Gates",* International Journal of Engineering Science and Technology, Vol. 2(8), 2010, 3838-3845.

[4] B. Murali Krishna, B. Raghu Kanth, " *A Comparative Study of Reversible Logic Gates",* International Journal of VLSI and Signal Processing Appications, Vol 2, Issue 1, Feb 2012, pp 51-55.

[5] Alexander Fish, Arkadiy Morgenshtein "*Gate-Diffusion Input(GDI) – A novel Power Efficient Method for Digital Circuits*", IEEE 2001: pp. 39-43.

[6] Li Qiao, You Yuxin, Wang Jinxiang and Ye Yizheng, "*VLSI Implementation of a High Speed and Low Power punctured Viterbi Decoder*", International Conference on Computers, Communications, Control and Power Engineering, IEEE transactions ,October 2002.

[7] C.Arun, V Rajamani, " *Design and VLSI Implementation of a Low Probablity of Error Viterbi Decoder",* First International Conference on Emerging Trends in Engineering and Technology, ICETET, July 2008.

[8] D. Chakraborty, P. Raha, A. Bhattacharya, *"Speed Optimization of a FPGA Based Modified Viterbi Decoder",* International Conference on Computer Communication and Informatics (ICCCI), January 2013.

[9] Landauer.R, "*Irreversibility and Heat Generation in the Computing Process*", IBM J. Research and Development: pp. 183-191, 1961.

[10] Song Li, Qing-Ming Yi, "*The Design of High-Speed and Low Power Consumption Bidirection Viterbi Decoder",* International Conference on Machine Learning and Cybernetics, August 2006.

[11] Hao Yang and Xiao-Lang Yan," *Design and Implementation of a High-Speed and Area-Efficient Viterbi Decoder",* 8th International Conference on Solid-State and Integrated Circuit Technology, ICSICT October 2006.

[12] Umar M Karim, Muhammad U.K Khan, Yahya M. Khawaja, "*An Area Reduced, Speed Optimized Implementation on Viterbi Decoder",* International Conference on Computer Networks and Information Technology (ICCNIT), July 2011.

[13] R.Tessier, S. Swaminathan, R. Rangaswamy, "*A Reconfigurable, Power-Efficient AdaptiveViterbi Decoeder",* IEEE Transactions on Very Large Scale Integration Systems, April 2005.

[14] F.Chan and D.Haccoun, *" Adaptive Viterbi Decoding of Convolutional Codes Over Memoryless Channels",* IEEE Transactions on Communications, 45(11): 1389-1400, November 1997.

[15] Kai-Ting Shr, Hong-Du Chen, Yuan-Hao Huang, " *A Low Complexity Viterbi Decoder for Space-Time Trellis Codes"*, IEEE Transactions on Circuits and Systems, April 2010.

[16] Chun-Yuan Chu, Yu-Chuan Huang, An-Yeu Wu," *Power Efficient Low Latency Survivor Memory Architecture for Viterbi Decoder",* IEEE Transactions 2008.

International Conference on Latest Trends in Electronics and Communication (ICLTEC)-ISBN 978-93-88808-62-0

DESIGN OF HIGH SPEED SEQUENCE DETECTOR USING VERILOG

Bejagam Divya[1]
M.Tech Scholar[1]
Department of Electronics and Communication Engineering[1]
Malla Reddy College of Engineering, Hyderabad, India[1]
bejagamdivya84@gmail.com[1]

Abstract: There is a enormous usage of sequence detectors in digital circuits as it is the basic function and it became essential in most of the digital systems counting ALU, microprocessors and DSP. Sequential circuit's works on a clock cycle which may be synchronous or asynchronous. Sequential circuits use current inputs and previous inputs by storing the information and putting back into the circuit on the next clock cycle. This paper presents the high speed Sequence Detector in Verilog, which is a sequential state machine used to detect consecutive bits in a binary string. The flip-flops help to detect the pattern in the given string. The Sequence Detector gives for some particular sequence of inputs and outputs, whenever the desired sequence has found. And this paper shows a great vision on the design analysis of sequence detector using Verilog. The delay (1.045ns) minimized. The proposed architecture of sequence detector is synthesized in Xilinx ISE14.7.

Keywords: Sequence detector, sequential circuits, flip-flop, delay, Verilog, Mealy machine and Xilinx ISE14.7.

I. Introduction

FSM is a mathematical model of computation. It is an abstract machine that can be in exactly one of a finite number of states at any given time. The FSM can change from one state to another in response to some external inputs; the change from one state to another is called a transition. An FSM is defined by a list of its states, its initial state, and the conditions for each transition. Finite state machines are of two types - deterministic finite state machines and non-deterministic finite state machines. A deterministic finite-state machine can be constructed equivalent to any non-deterministic one.

The behavior of state machines can be observed in many devices in modern society that perform a predetermined sequence of actions depending on a sequence of events with which they are presented. Simple examples are vending machines, which dispense products when the proper combination of coins is deposited, elevators, whose sequence of stops is determined by the floors requested by riders, traffic lights, which change sequence when cars are waiting, and combination locks, which require the input of combination numbers in the proper order. The finite state machine has less computational power than some other models of computation such as the Turing machine.

II. Proposed work

A sequence detector is a sequential state machine. Sequential circuit's works with respect to a clock cycle which may be synchronous or asynchronous. The figure shows a basic diagram block of sequence detector. Sequential circuits use current inputs and previous inputs by storing the data and putting back into the circuit on the next clock cycle.

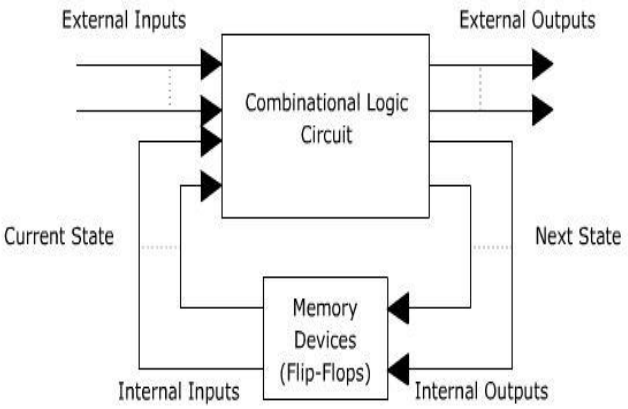


Figure 1: Block Diagram of Sequence Detector.

Finite State Machine (FSM):

A FSM is a model it is used to design sequential logic circuits. It is considered as an abstract machine that can be in one of a finite number of states. The machine is accessible in only one state at a time; the state it is in at any given time is called the current state. It can change

from one state to another state when the triggering event or condition is introduced into the machine, this is called a transition. A specific FSM is well-defined by a list of its states, and the triggering condition for each transition. It can be implemented using models like Mealy machine and Moore machine. For this expt., we will use Mealy machine model implementation.

Types Of Sequence Detector:
There exist basically two types of sequence detectors.
That are:
1) Overlapping Sequence Detector.
2) Non- Overlapping Sequence Detector.

Overlapping Sequence Detector:
In a sequence detector that allows overlap, the final bits of one sequence can be the start of another sequence.

Non-Overlapping Sequence Detector:
The sequence detector with no overlap allowed resets itself to the start state when the sequence has been detected.

Flip-flop: A flip-flop or latch is a circuit that has two stable states and it can be used to store state data. A flip-flop is a bistable multivibrator. The circuit can be made to change state by signals applied to one or more Control inputs and will have one or two outputs, one for the normal value and one for the complement value of the stored bit. Memory elements in any sequential circuit are usually flip-flops.

Sequence detector:
Suppose a sequence detector is to be designed to detect a sequence 1101.
Then the state diagram will be:

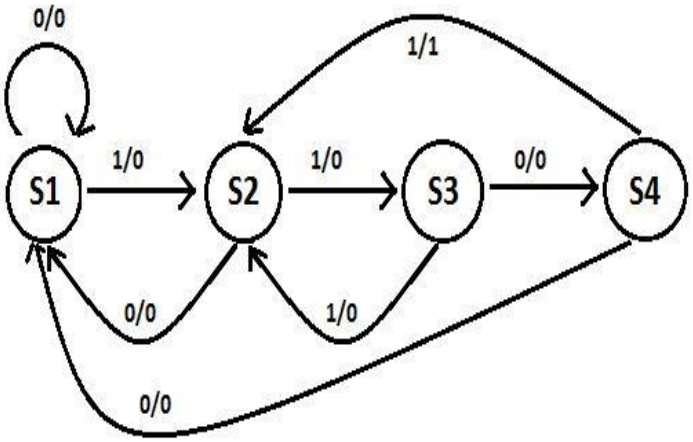


Figure 2: State Diagram of Sequence Detector.

Note that this state diagram is considering overlap i.e. if we have input 1101101 we will have output0001001. For same input, non-overlap case will have output 0001000. Either cases are correct but we will consider only overlap case henceforth.

III. Synthesis Results

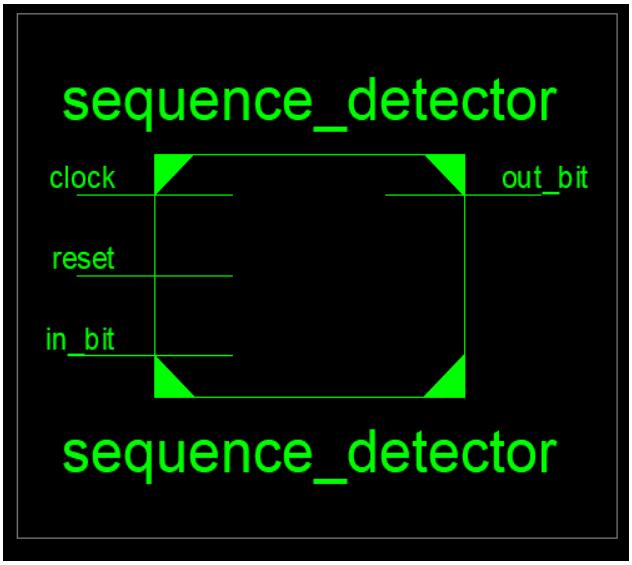


Figure 3 : Top Module of Sequence Detector.

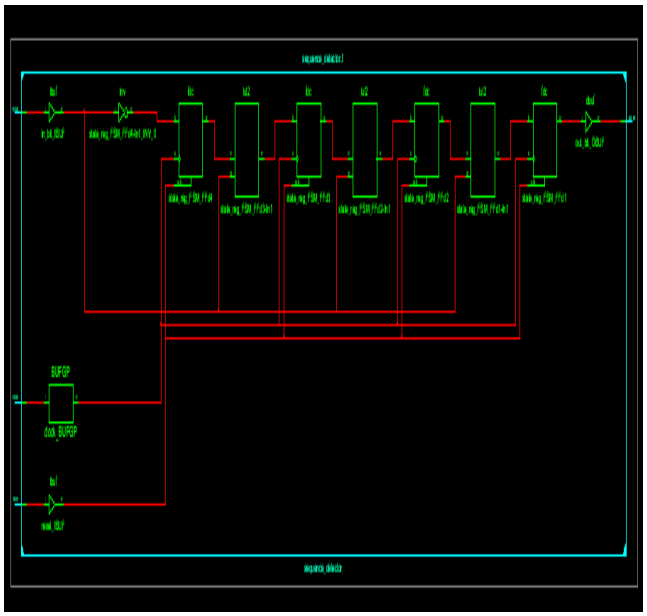


Figure 4 : Technology Schematic of Sequence Detector.

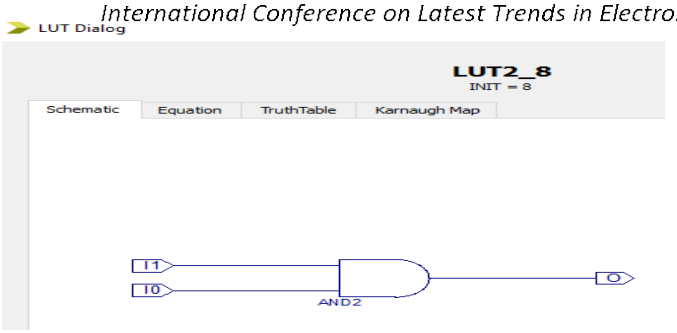


Figure 5 : LUT Schematic of Sequence Detector.

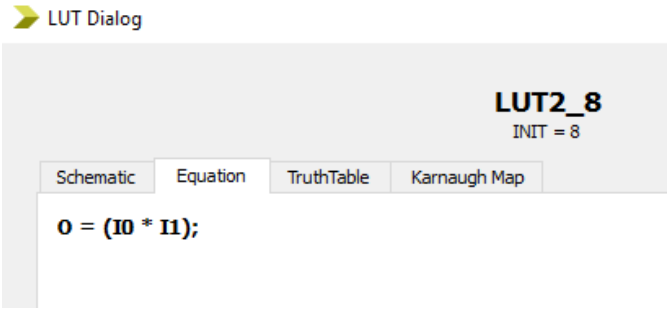
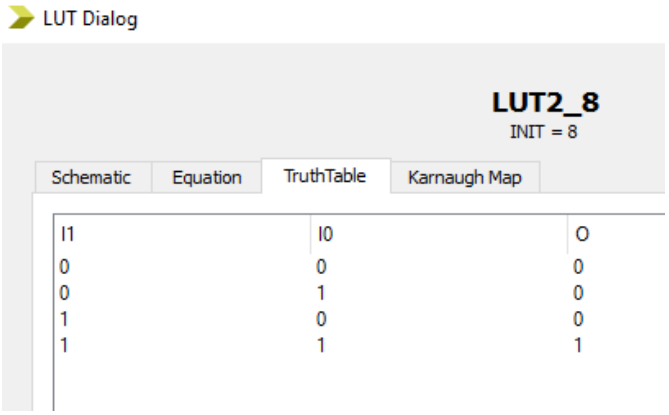


Figure 6 : LUT Equation of Sequence Detector.



I1	I0	O
0	0	0
0	1	0
1	0	0
1	1	1

Figure 7 : LUT Truth table of Sequence Detector.

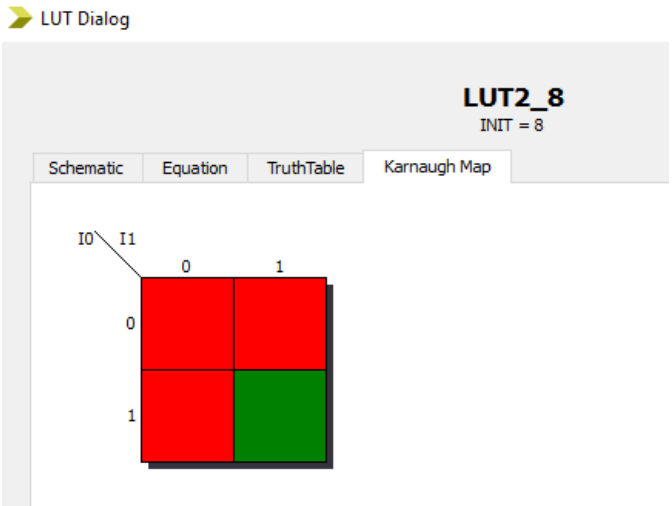


Figure 8 : LUT karnaugh map of Sequence Detector.

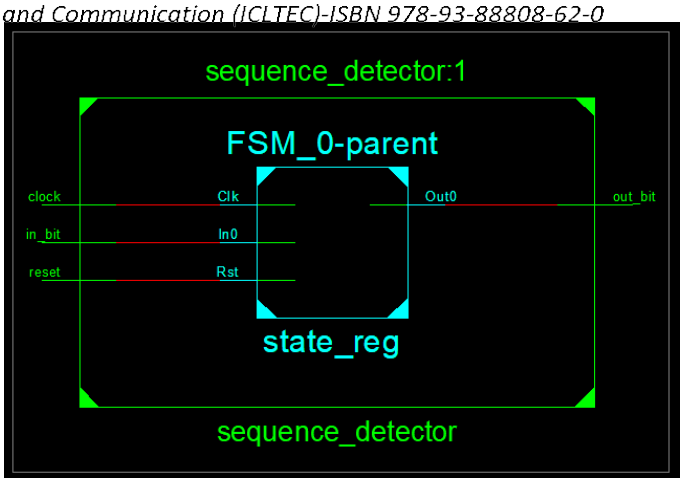


Figure 9 : Top RTL Schematic of Sequence Detector.

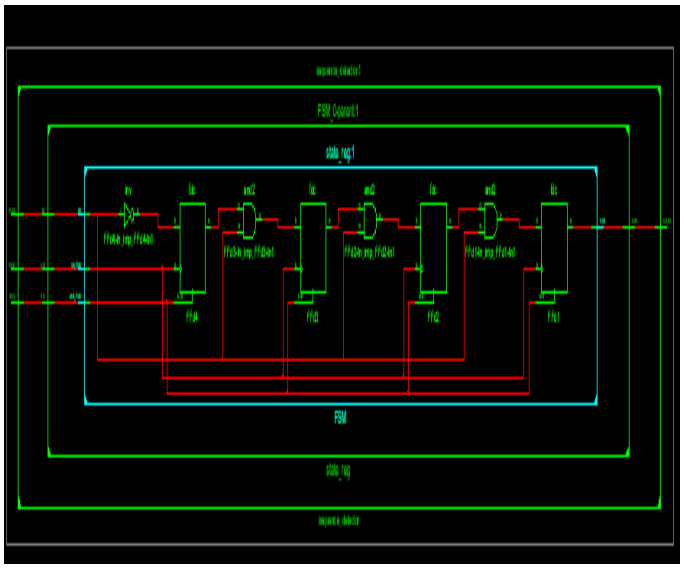


Figure 10 : RTL Schematic of Sequence Detector.

Device Utilization Summary (estimated values)			[-]
Logic Utilization	Used	Available	Utilization
Number of Slice Registers	4	44800	0%
Number of Slice LUTs	4	44800	0%
Number of fully used LUT-FF pairs	0	8	0%
Number of bonded IOBs	4	640	0%
Number of BUFG/BUFGCTRLs	1	32	3%

Table 1: Device Utilization of Sequence Detector.

Final Register Report:

Macro Statistics

# Registers	: 4
Flip-Flops	: 4

Cell Usage :

# BELS	: 4
# INV	: 1
# LUT2	: 3
# FlipFlops/Latches	: 4
# FDC	: 4
# Clock Buffers	: 1
# BUFGP	: 1
# IO Buffers	: 3
# IBUF	: 2
# OBUF	: 1

Timing Summary:

Speed Grade: -1

Minimum period: 1.045ns (Maximum Frequency: 956.938MHz)

Minimum input arrival time before clock: 1.744ns

Maximum output required time after clock: 3.259ns

Timing Detail:

Timing constraint: Default period analysis for Clock 'clock'

Clock period: 1.045ns (frequency: 956.938MHz)

Total number of paths / destination ports: 3 / 3

Delay: 1.045ns (Levels of Logic = 1)

Source: state_reg_FSM_FFd2 (FF)

Destination: state_reg_FSM_FFd1 (FF)

Source Clock: clock rising

Destination Clock: clock rising

Total 1.045ns (0.565ns logic, 0.480ns route)
(54.1% logic, 45.9% route)

Total memory usage is **4568248 kilobytes**

IV. Simulation result

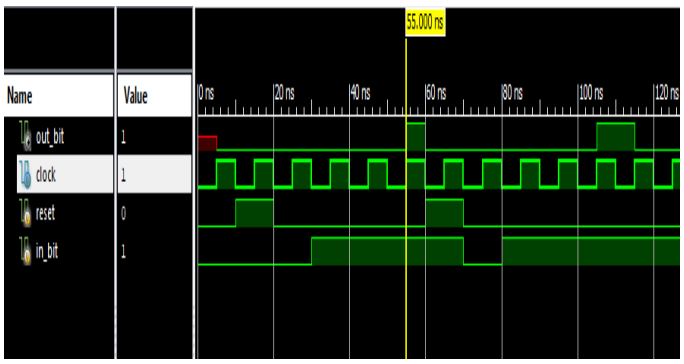


Figure 11: Simulation Result of the Sequence Detector.

The results are obtained by simulating the verilog code in Xilinx ISE 14.7.

V. Conclusion

The proposed design of sequence detector is synthesized and simulated in Xilinx ISE 14.7. The source code is written in Verilog. As we know Delay is the major factor in VLSI design that limits the performance of any circuit. This paper concentrates more on speed by presenting a simple approach to reduce the delay of sequence detector architecture, which helps in increasing the computational level of calculations. This proposed sequence detector has delay 1.045ns. Sequence detector has extensive variety of applications such as design of Ring counter, Serial Adder, Schmitt trigger.

References

1. IEEE Standard 1364-2001, IEEE Standard Description Language based on the Verilog Hardware Description Language, 2001.
2. Mindek, M., “Finite State Automata and Image Recognition” DATESO 2004, pp 132-143 (2004), ISBN: 80-248-0457-3.
3. G. Navarro, R. Baeza-Yates, “Improving an Algorithm for Approximate String Matching.”,Algorithmica, 30(4) 2001.
4. S.V.Ramasubramanian and Kamala Krithivasan,“Finite Automata and Digital Images”, IJPRAI, Vol. 14, No. 4, pp. 501-524, 2000.
5. John C.Martin , ”Introduction to Languages and Theory of Computation”,Third edition, Tata McGraw Hill,2009.

A low power large bandwidth flipped voltage follower based current mirror

P.Anil kumar1 **Rajesh Durgam2**
***1anilpole403@gmail.com , *2 rajeshdurgam2000@gmail.com**
1,2 Assistant Professor,Department of ECE, Malla Reddy College of Engineering, Hyderabad, Telangana, INDIA

Abstract
In this study, simple nMOS current mirror has been modified and the performance has been improved employing high performance flipped voltage follower (FVF). The resistively compensated FVF has been utilized at the input side of the current mirror. The advantages of using flipped voltage follower include high linearity, wide bandwidth and reduced power dissipation.

In this study, the simulation responses of all the circuits are presented. The functionality and performance improvement of all the circuits are simulated on Spectre simulator (Cadence) using model parameters of TSMC 0.18 μm CMOS BSIM 3 and level 49 technology.

Keywords: Current mirror, low power, analog circuit, cascode, voltage follower.

Introduction

The explosive growth in electronics world towards portability and high-speed VLSI systems has motivated the current research in the direction of high frequency analog basic cells. The requirement of high-performance analog devices in the communication systems has increased the usage of current-mode circuits. Current mirror is among the most important and essential current mode device that has been used in numerous analog systems.[1-9]

The important features of a current mirror include accurate current mirroring, large input and output current swings, high output impedance and good linearity. In the design of a current mirror the key issues are the improvement of high-frequency characteristic and the realization of high output impedance. Also, a current mirror which has high power consumption and small -3 dB frequency is not reliable for low-voltage high-speed applications. Several large bandwidth current mirrors are reported in literature.[1]

Gupta et al[10] have improved the frequency performance of FVF based current mirror by introducing both passive[3,4] and active resistance[3,4] at the gate of input transistor pair of the low voltage current mirror in. An approximately 200 MHz improvement in the bandwidth of passively and actively compensated current mirrors has been obtained.[10] In this work, the conventional current mirrors (CMs) have been redesigned for high frequency applications.

Simple Current Mirror

Voo and Toumazou[3] have improved the bandwidth of the simple current mirrors by using resistive compensation technique. They have introduced a compensating resistor in between the drain and gate of the input MOS transistor of the current mirror. This leads to introduction of one-pole and one-zero in the transfer function and the zero is used to cancel the dominant pole (pole-zero cancellation). Hence, the bandwidth of the system will get increased.[3] A simple CM consists of 2 MOS transistors with compensating resistor R as shown in fig. 1.

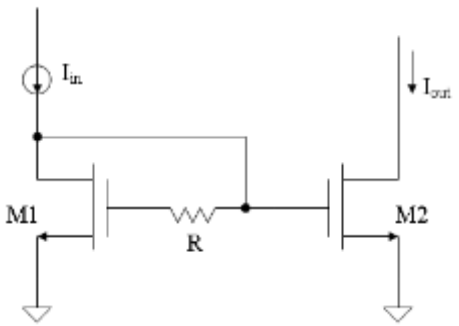


Fig. 1: Simple CM [3]

The transfer function (TF) of the current mirror without compensating resistor R[3] is:

$$A_i(s) = \frac{g_{m2}}{g_{m1}} \frac{1}{\left(1 + s\left(2C_{gs}/g_{m1}\right)\right)}$$ and bandwidth is:

$$\omega_0 = \left(\frac{g_{m1}}{2C_{gs}}\right)$$

where g_m is the transconductance and C_{gs} is the gate-source capacitance of each MOS transistor. With compensating resistor, TF of the current mirror is transformed:

$$A_i(s) = \frac{g_{m2}}{C_{gs2}} \frac{\left(s + 1/RC_{gs1}\right)}{\left(s^2 + \left(\frac{C_{gs1} + C_{gs2}}{RC_{gs1}C_{gs2}}\right)s + \left(\frac{g_{m1}}{RC_{gs1}C_{gs2}}\right)\right)} \quad (1)$$

The transfer function of the simple current mirror is transformed from first order single pole to second order low pass consisting of 1 zero and 2 poles. The zero and poles of the resistively compensated simple current mirror are:

$$Z_1 = -\frac{1}{RC_{gs1}} \quad \text{and}$$

$$P_{1,2} = \frac{C_{gs1}+C_{gs2}}{2RC_{gs1}C_{gs2}}\left[-1 \pm \sqrt{1-\frac{4g_{m1}RC_{gs1}C_{gs2}}{\left(C_{gs1}+C_{gs2}\right)^2}}\right] \quad (2)$$

From equation (1), the bandwidth of the system is obtained as (3):

$$\omega_0 = \sqrt{\frac{g_{m1}}{RC_{gs1}C_{gs2}}} \quad (3)$$

For $R = 1/g_{m1}$ and $C_{gs1} = C_{gs2}$, the zero gets cancelled with one of the poles resulting into a first order transfer function. The -3 dB frequency of the compensated CM[3] is: $\omega_0 = \left(\frac{g_{m1}}{C_{gs}}\right)$. It can be concluded that the -3 dB frequency of the resistively compensated CM is twice that of the previous one. Same method has been further applied for bandwidth increment of a FVF based cascode current mirror.

The conventional LVCCM

The conventional simple current mirror structure has drawbacks of low ratio of output to input impedance. Some circuits were reported earlier such as regular cascode current mirrors to improve the output impedance but suffer from increased minimum supply voltage which limited the applicability of these structures for low voltage operation.[11-14] In order to meet the present electronics industry requirements of low power supply, many circuits are available and the most commonly used cell is the FVF based LVCCM.[1,15-19] It can is seen that the performance of FVF based cascode current-mirrors (shown in fig. 2) including maximum operating signal and error of current transfer, is better in comparison to conventional LVCCM.

Thus, FVF based CMs could be used in good-performance and low operating voltage analog systems. If all transistors of the current mirror are in the saturation region, shunt feedback causes impedance at input node to be low. Thus, the current flow amount through this input node will not affect its voltage. Hence designer can achieve high performance current mirror.[10] Another CM topology which is extensively used in analog application is alternatively-fed FVF based cascode current mirror[1] as depicted in fig. 3.

The minimum required supply voltage (V_{DD}) and minimum output voltages (V_{out}) are expressed:[1,20]

$$V_{DD,\min} = V_{TH(Mn11)} + V_{DS,sat(Mp1)} \quad (4)$$

$$V_{in,\min} = V_{TH(Mn11)} + V_{DS(Mn5)} \quad (5)$$

$$V_{out,min} = V_{DS,sat(Mn2)} + V_{DS,sat(Mn3)} \quad (6)$$

In order to further decrement of the input resistance (R_{in}), the topology of FVF based cascode CM is modified in a way that input current is fed at a different node i.e. output node of the input FVF [1]. The minimum required V_{DD} and minimum V_{out} are obtained as: [1,20]

$$V_{DD,min} = V_{TH(Mn1)} + V_{DS,sat(Mp1)} \text{ and } V_{in,\min} = V_{DS,sat(Mn11)}.$$

It can be noticed from equations (9) and (12) that there is a significant reduction in input voltage. The Rin of the FVF based cascode CM is given by the expression:

$$R_{in} \cong \frac{2}{g_{m1}g_{m5}r_{o5}}$$

where gmi (i=1, 5) is the transconductance of Mi and r_{o5} is the output impedance of the M5 transistor.

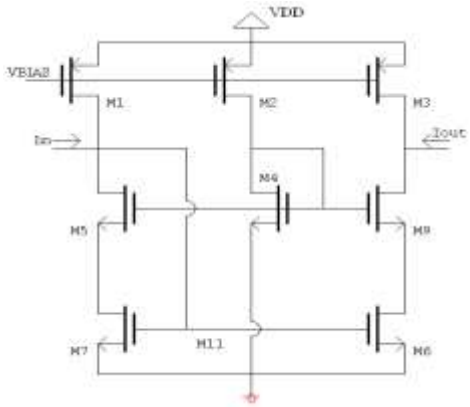


Fig. 2: FVF based LVCCM[1]

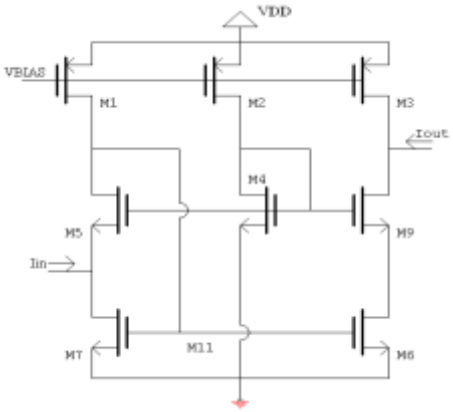


Fig. 3: Alternatively fed FVF based LVCCM

Resistively compensated FVF based LVCCM

The resistively compensated low voltage FVF based CM has been designed. The wideband flipped voltage follower has been inserted at the input terminal of the LVCCM shown in fig. 4 to enhance the bandwidth of the circuit.[10] The modified CM is shown in fig. 5 (R_{COMP} and R are compensating and feedback resistance respectively). Fig. 6 shows the actively compensated CM suggested by Gupta et al[10] and the modified version of fig. 6 is shown in fig. 7 (M_{COMP} is the transistor used for active compensation).

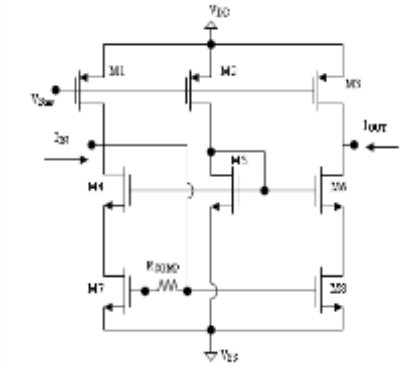


Fig. 4: Passively compensated CM[10]

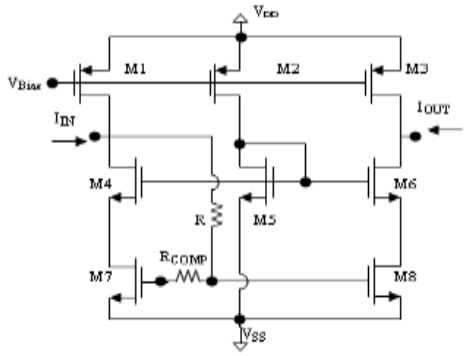


Fig. 5: Modified CM with the resistively compensated FVF

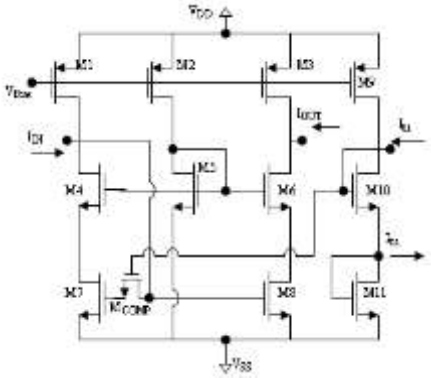


Fig. 6: Actively compensated CM[10]

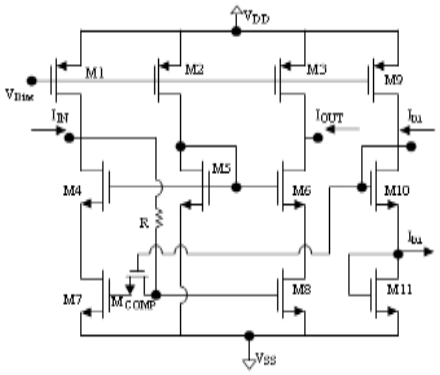


Fig. 7: Modified actively compensated CM with the resistively compensated FVF

Simulation Results and Discussion

Spectre simulator of Cadence using model parameters of TSMC 0.18 μm CMOS BSIM 3 and level 49 technology has been used to authenticate the functionality and performance development of all analog circuits. The simulation results of all the compensated LVCCMs are shown here.

The error (Iout-Iin) is shown in fig. 8 and it is almost -3.37%. The input and output compliances are portrayed in figures 9 and 10 respectively. From fig. 11, it can be seen that at 50 μA input current the modified CM dissipates 324.3 μW. It can be seen that when CM circuits are modified, the DC performance factors do not vary. Fig. 12 and 13 show the effect of resistive compensation on both the resistances of CMs. The obtained input and output resistances of CMs are 1.935 kΩ and 0.22 MΩ respectively. It can be observed from fig. 12 that the Rin of the CM decreases with frequency as the value of feedback resistance R increases (conventional CM is shown by solid line, CM with R = 1.6 kΩ and 6 kΩ are shown by dotted line with cross marker and dashed line with circle marker respectively).

Therefore, it leads to enhanced current flow at node output of FVF and input node of CM. From fig. 13, it can be seen that the output resistance of the designed LVCCM is same. Fig. 14 shows the frequency responses of CMs (fig. 2 and fig. 4 and 5). An improvement of 1.2 GHz in -3dB frequency is achieved by using the wideband FVF in passively compensated CM[10] i.e. BWER is 1.3 approximately.

However, peaking has been observed to achieve BWER of 1.6 which limits the maximum value of the compensating resistor. The frequency responses of the conventional and modified (fig. 6 and 7) actively compensated CMs can be depicted in fig. 14. The BWER is 1.2 of the improved actively compensated CM. It can be seen from figures 14 and 15 that the bandwidths of the passively compensated CMs are larger than that of the actively compensated CM. Active implementation of the compensating resistor provides several advantages such as smaller chip area requirement, but it provides smaller bandwidth than a passive resistor and increases peaking in the frequency response.

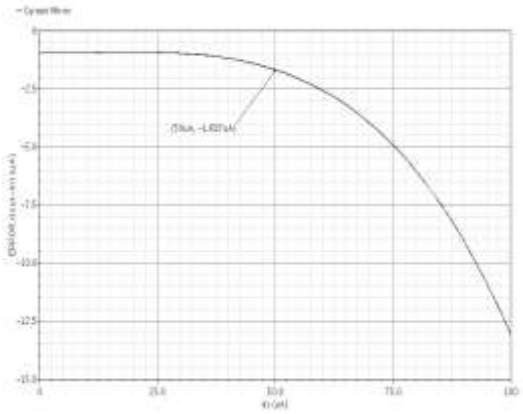


Fig. 8: Transfer error (Iout-Iin)

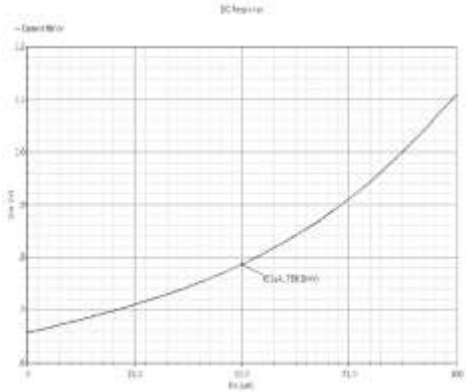


Fig. 9: Input Voltage (Vin) Versus Input Current (Iin)

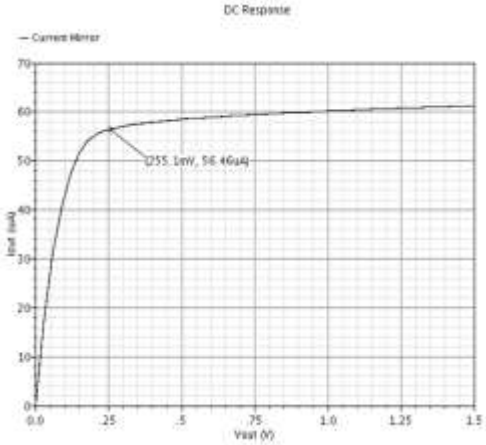


Fig. 10: Output Current (Iout) as a function of Vout

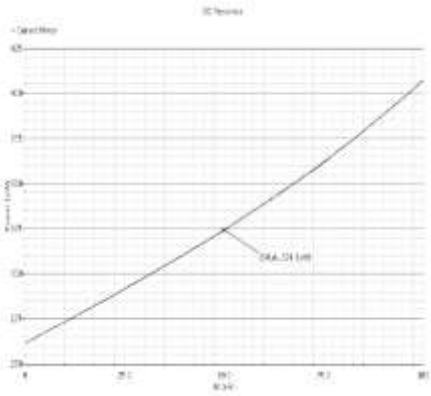


Fig. 11: DC power consumption

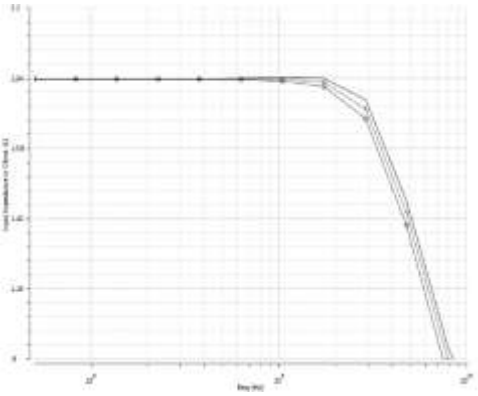


Fig. 12: Rin of (Fig. 2) and (Fig. 4 and 5) CMs

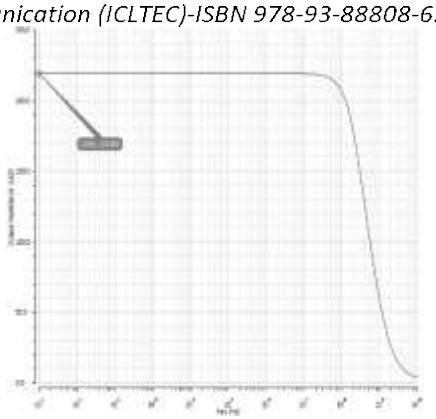


Fig. 13: Output impedance w.r.t frequency

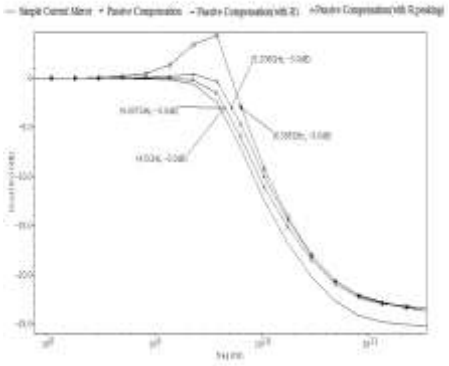


Fig. 14: Frequency responses of passively compensated CM (Fig. 5)

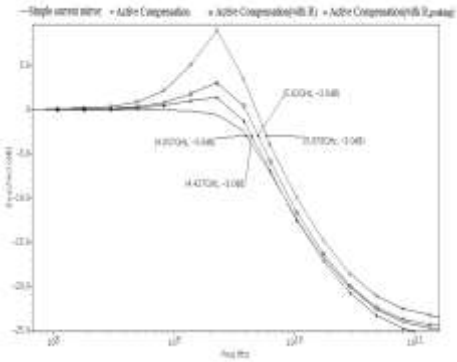


Fig. 15: Frequency responses of actively compensated CM (Fig. 7)

Conclusion

This work is dedicated to the development of FVF based conventional CM with lower power consumption. A resistively compensated FVF is used in place of conventional one in the low voltage CM to increase the -3dB frequency. The bandwidth of the proposed wideband passively compensated and actively compensated CMs is 6.395 GHz and 5.02 GHz respectively. From simulation results it has been inferred that the designed CMs circuits exhibit large bandwidth without any variation in the DC performances and hence, these circuits may find wide range of applications in high speed signal processing systems.

References

1. Koliopoulos C. and Psychalinos C., A Comparative study of the performance of the flipped voltage follower based low voltage current mirror, IEEE Int. Symp. Circuits and Systems, 1-4 **(2007)**

2. Carvajal R., Ramirez-Angulo J., Lopez Martin A., Torralba A., Galan J., Carlosena A. and Chavero F., The Flipped voltage follower: a useful cell for low voltage low power circuit design, *IEEE Trans. on Circuits and Systems*, **52(7)**, 1276 – 1279 **(2005)**

3. Voo T. and Toumazou C., High-speed current mirror resistive compensation technique, *IEE Electronics Letters*, **31(4)**, 248 – 250 **(1995)**

4. Voo T. and Toumazou C., Precision temperature stabilised tunable CMOS current-mirror for filter applications, *IEE Electronics Letters*, **32(2)**, 105-106 **(1996)**

5. Voo T. and Toumazou C., Efficient tunable continuous-time integrated current-mode filter designs, IEEE International Symposium on Circuits and Systems, 93-96 **(1996)**

6. Rajput S.S. and Jamuar S.S., A high performance current mirror for low voltage design. IEEE Circuits and Systems, In proc IEEE APCCAS, Tianjin, China 170-173 **(2000)**

7. Rajput S.S. and Jamuar S.S., Advanced current mirrors for low voltage analog designs, Proc. of IEEE ICSE '04, Kuala Lumpur, Malaysia, 258-263 **(2004)**

8. Rajput S.S. and Jamuar S.S., Low voltage, low power, high performance current mirror for portable analogue and mixed mode applications. Circuits, Devices and Systems, IEE Proc., 273 – 278 **(2001)**

9. Jamuar S.S. and Rajput S.S., Low-voltage current-mode analog circuit structures and their applications, IEEE International ASIC/SOC Conf., 477- 478 **(2002)**

10. Gupta M., Aggarwal P., Singh P. and Jindal N.K., Low voltage current mirrors with enhanced bandwidth, *Analog Integrated Circuits Signal Process*, **59**, 97-103 **(2009)**

11. Prodanov V. and Green M., CMOS current mirrors with reduced input and output voltage requirements, *IEE Electronic Letters*, **32(2)**, 104-105 **(1996)**

12. Ledesma F., Garcia R. and Ramirez-Angulo J., Comparison of new and conventional low voltage current mirrors, Midwest Symp. on Circuits and Systems, 49-52 **(2002)**

13. Yonghua C. and Geiger R.L., Cascode current mirrors with low input, output and supply voltage requirements, IEEE Midwest Symp. on Circuits and Systems, 490 – 493 **(2000)**

14. Palumbo G., On the high-frequency response of CMOS cascode current mirror, IEEE proc., 7th Mediterranean Electrotechnical Conf., 1211-1214 **(1994)**

15. Ramirez-Angulo J., Gupta S., Padilla I., Carvajal R.G., Torralba A., Jimenez M. and Munoz F., Comparison of conventional and new flipped voltage structures with increased input/output signal swing and current sourcing/sinking capabilities, *Midwest Symposium on Circuits and Systems*, 1151-1154 **(2005)**

16. Ramirez-Angulo J., Carvajal R.G., Torralba A. and Galan J., Low supply voltage high-performance CMOS current mirror with low input and output voltage requirements, *IEEE Trans. on Circuits and Systems-II*, **51(3)**, 124-129 **(2004)**

17. Meaamar A., Low-Voltage, High-Performance Current Mirror Circuit Techniques, IEEE International Conf. on Semiconductor Devices, Kuala Lumpur, Malaysia, 661-665 **(2006)**

18. Torralba A., Carvajal R.G., Ramirez-Angulo J. and Munoz E., Output stage for low supply voltage, high-performance CMOS current mirrors, *Electronics Letters*, **38(24)**, 1528 – 1529 **(2002)**

19. Aggarwal B., Gupta M. and Gupta A.K., A comparative study of various current mirror configurations: Topologies and characteristics, *Microelectronics Journal*, **53**, 134-155 **(2016)**

20. Razavi B., Design of Analog CMOS Integrated Circuits, McGraw-Hill, Inc. **(2001)**.

The Gas Leak Detection Based on a Wireless Monitoring System

K.SUHASINI, MTech scholor, Department of Electronics and Communication Engineering,

Hyderabad, India. suhasinikangati@gmail.com

1

***Abstract*—The industrial gas leaks cause accidents and pose threats to the environment and human life. Thus it is essential to detect the gas leaks in time. Usually, the abnormal concentration signals are defined by a fixed concentration value, such as 25% of the lower explosive limit (LEL). However, it is difficult to accumulate to the fixed point quickly when the leak is small. In addition, the actual leak signals are seldom available, making many data classification inoperable. To solve these problems, this paper proposes a detection approach by using the auto-correlation function (ACF) of the normal concentration segment. The feature of each normal segment is obtained by calculating the correlation coefficients between ACFs. According to the features of statistical analysis, a non-concentration threshold is determined to detect the real time signals. In addition, the weighted fusion algorithm based on the distance between the sensors and virtual leak source (VLS) is used to fuse multi-sensory data. The proposed method has been implemented in a field by building a wireless sensor network (WSN). It is confirmed that the system detection rate reaches as high as 96.7% and the average detection time delay is less than 30s on the premise of low false alarm rate.**

***Index Terms*—Gas Leak Detection, the Auto-correlation Function (ACF), the Correlation Coefficient, the Weighted Fusion, Wireless Sensor Network (WSN).**

I.INTRODUCTION

Globally, the gas, petroleum, chemical, metallurgy and other industries produce a large number of flammable and toxic gases as well as benzene and other organic vapors every year. Due to improper man made operation or equipment aging, a large number of gas leaks have occurred. Some leaks that are not detected or repaired in time have caused great economic losses, environmental pollution and even huge casualties. Statistically, 778 safety accidents occurred in China from January 2016 to January 2017. The explosions and poisonings resulted from the gas leakage accounted for 6.17% and 5.4% reference to the total accidents, respectively [1]. The usages of hazardous materials are strictly restricted for the safety by the State Administration of Work Safety. However, the usage of chemical and hazardous materials increases year by year. At present, a large number of commercial companies and research institutions engage in researching the detection of dangerous gas leak. In the market, the handheld devices are most commonly used. These devices have high accuracy and accurate positioning. However, the users need to go to the detected areas where the leaks are prone to occur while their safety and real time detection can't be ensured [2]. The mobile robot with vision and gas sensors can replace the workers to detect gas leaks and send the data to remote control center that is a novel detection method and protects the lives of workers [3]. However, this detection system is limited by the mobility of the robot and is not suitable for the monitoring area with uneven terrain and wide range. In addition, the wire monitoring system that is composed of high resolution sensors and control centers solves the problems of the security and real time issues by using the cables to transfer data and compromises on the inflexible installation and expensive maintenance [4].

Recently, the sensor technology, wireless communication and the embedded technology are developing rapidly. The wireless sensor network (WSN) with low cost, flexible installation, real-time continuous detection and other advantages is used widely in the monitoring of environmental parameter, structural health [5] and gas leak. Somov et al. [6] deployed a ZigBee WSN to monitor the gas leaks in the boiler facility. The system consists of 9 battery powered wireless sensor nodes and 1 network ZigBee coordinator while the system also has access to the Ethernet and GSM network to send messages. Generally, the system of detecting gas leak by gas sensors sets a few fixed leak concentration to confirm gas leak or not. Jelicic et al. presented a wireless sensor network (WSN) for monitoring indoor air quality. The sensor node is designed with very low sleep current consumption and the network is multimodal. It exploits information from auxiliary sensors and neighbor nodes about gas concentration to modify the behavior of the node and the measuring frequency of the gas concentration.[7] Rossi et al. exploited the transient response of the sensing element, so that indoor air quality assessment power required is

This work is supported in part by key research and development plan project of Zhejiang province (No.2018C01036), the National Natural Science Foundation of China (No. 61871167 and 61376117). (Corresponding author: Linxi Dong and Gaofeng Wang)

Linxi Dong is with the School of Electronics and Information, Hangzhou Dianzi University, Hangzhou 310018, China, the State Key Lab of Functional Materials for Informatics, Shanghai Institute of Microsystems and Information Technology, CAS, Shanghai 200050, China, and also with the State Key Lab for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China (Email: donglinxi@hdu.edu.cn). Libo Zhao is with State Key Laboratory for Manufacturing Systems Engineering, School of Mechanical Engineering in Xi'an Jiaotong University , China(Email: libozhao@xjtu.edu.cn). Zhiyuan Qiao, Haonan Wang, Weihuang Yang, Wensheng Zhao, Kuiwen Xu, Gaofeng Wang are all with the Key Laboratory of RF Circuits and System of Ministry of Education, Electronic and Information College of Hangzhou Dianzi University, Hangzhou 310018, China. (Email: 2569173072@qq.com,1142214134@qq.com,89297508@qq.com,wshzhao@hdu.edu.cn,kuiwenxu@hdu.edu.cn,gaofeng@hdu.edu.cn). Haixia Yan is with School of Information Engineering, Hangzhou Dianzi University, Hangzhou 310018, China (Email: 1328515466@qq.com)

reduced to 1/20 of the original.[8] In the paper [9], two alert levels that are high alert and low alert are set respectively. According to the setting specifications for safety monitoring and controlling equipment in dangerous chemicals major hazard installations, two fixed points are set at 25% of the lower explosive limit (LEL) and 50% LEL usually. However, the leaks cannot be detected by setting a fixed concentration threshold generally when the leakage is small or just starts to happen.

Different from the common detection method of gas concentration for the leaks, the methods of infrared and ultrasonic detection develop rapidly in recent years. The former method uses thermal imaging and infrared image processing to detect leaks [10]. For example, a low cost infrared camera called SENSIA's Gas Imaging System [11] developed by Universidad Carlos III de Madrid is a low cost infrared camera based on spectrally adapted and high sensitivity imaging technology. It can easily detect and identify fugitive gas emissions. Additionally, Leis *et al.* proposed improvements in the detection time [12] and the optical flux variation due to using solid-state IR sources to heat [13] respectively, which promoted the development of infrared sensors in leak detection. The ultrasonic method detects the leaks by analyzing the ultrasound that generated under pressure. The approach is not affected by inclement weather, wind direction, and can respond quickly without physical contacts. The GDU-Incus ultrasonic gas leak detector released by the United States Emerson Group can detect the gas leak in the range from 2 to 40 meters [14]. Also, it is unnecessary to wait until the gas concentration accumulates to the dangerous threshold. In the actual environment, the two kinds of devices are often set on the towers in order to monitor a larger area. Both of them are difficult to detect the leaks with low leakage pressure, the environmental obstacles, or interferences of heat sources.

In order to solve the small leak detection, many algorithms have used in the field of pipeline leak detection. The probabilistic and data classification methods are the common approaches. Akouemo *et al.* [15] used a linear regression model and a geometric probability distribution of the residuals to determine the anomalous probability of a data, and then trained a Bayesian maximum likelihood classifier to distinguish between false positives and true anomalies. Gupta *et al.* [16] proposed a probabilistic method which used Bayesian probabilistic framework and the steady state flow and pressure values of gas to detect the presence of a leak event. In addition, Wang *et al.* [17] proposed a pipeline leak detection approach by using time-domain statistical features from normal sample signals, and built the support vector data description (SVDD) model to finish the leak detection. For the small leaks, some results have been achieved. Xiao *et al.* [18] used the variational mode decomposition (VMD) to do the components reconstruction and proposed the ambiguity correlation classification (ACC) based on the correlation coefficient to detect the small leak of pipeline. A novel de-noise algorithm based on dual tree complex wavelet transform and singular value decomposition (DTCWT-SVD) is applied for small leak detection [19]. Similarly, the harmonic wavelet based pipeline small leakage detection method is proposed by Hu *et al* [20]. Furthermore, Kang *et al.* [21] presented a water leakage monitoring architecture using the one dimensional convolutional neural network and a support vector machine (1D-CNN-SVM) that utilized the feature maps of CNN as input to the classifier and the approach does not need data transformation.

For the small leaks, it is difficult for using the concentration threshold distinguish the small leakage. So the features should to be extracted in the term of time-domain or frequency-domain so that the abnormal signals are distinguished. In the field of pathological analysis of electrocardiogram (ECG), an approach is proposed which extracted the Auto-correlation Functions (ACFs) of segments and analyzed the similarity between the ACFs of normal and abnormal to define the ECG signals [22]. The essence of this method is the thought of random signal processing [23]. However, the approach is to analyze whether the historical data contains anomalies after the sampling of signals have been completed. Therefore the approach cannot be directly used for the online leak detection. Pister *et al.* [24] proposed a method combined the thought in ECG and the likelihood function to detect the industrial gas leaks. By establishing the concentration distribution model, the likelihood function is obtained, and then the result of leak detection is got by the method mentioned in the ECG. However, the threshold is obtained through constantly trying, and the impact of sensor node's position was not considered.

To tackle these limitations, this work further advanced the previous method from the following perspectives:

- Implement an automatic procedure to select sensor nodes that participate in leak detection, considering the location of sensors;
- Explicitly define the degree of the auto-correlation function (ACF) as the feature for leak detection;
- Propose a general procedure for the threshold that distinguishes between abnormal and normal;

In the proposed method, normal data are sampled under the normal operation environment. The sample is first divided into fixed length segments. Then, the ACF of each segment is obtained and the degree of each segment is calculated as the time-domain feature by computing the correlation coefficient between each pair of ACF. It is employed as the discriminative feature for leak detection. Furthermore, the threshold, which is defined as the baseline to detect the abnormal signal, is obtained by analyzing the distribution of the degrees of the normal signals.

In this work, the data collected by sensor nodes firstly are initially analyzed and classified. The concentration is divided into three levels: the first level is greater than 50% LEL; the second level is within 25-50% LEL; the third level is less than 25%LEL. The first two cases are defined as concentration abnormal while the third level is undefined. In addition, the temperature, pressure and humidity are divided into normal and abnormal according to fixed alarm values, respectively. The system in this study will sound alarms immediately when the abnormal results are obtained. And the undefined concentration data will be defined by the method proposed in this paper. The reliability is verifed by comparing the measurement data of the professional sensor equipment with the data in our system.

The rest of the paper is organized as follows. The wireless monitoring system for the gas leak detection is introduced in Section II. Section III describes the method of the gas leak detection including the selection of the multisensory data and the feature extraction in time-domain mainly. The experimental

results are presented in Section IV. And this paper is summarized and concluded in Section V.

II.WSN Based Detection System

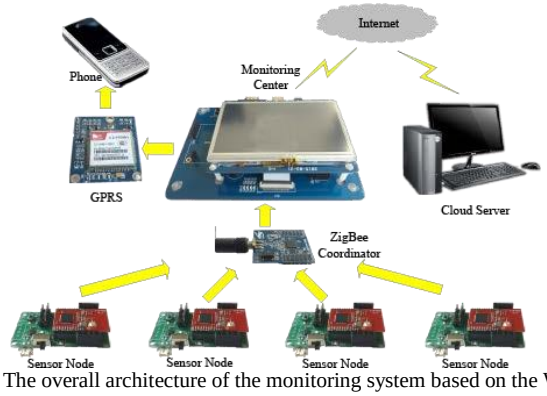


Fig. 1. The overall architecture of the monitoring system based on the WSN for the detection of gas leak.

Fig. 1 illustrates the overall architecture of the monitoring system based on the WSN for the detection of gas leak in an open environment. In order to detect the gas leak, a number of ZigBee nodes with gas sensors and a ZigBee coordinator need to be deployed to form a WSN in the monitored area. After converting and filtering, the data of gas concentration will be sent to the ZigBee coordinator. Then, the sensor nodes which send the third level data are selected to form a set by the ZigBee coordinator. The ZigBee network uses the 2.4GHz band. And the data of the sensor nodes in the set is fused into one concentration sequence (CS). Then the CS is transferred to the monitoring center of system and judged whether the gas leaks by the detection algorithm. The method used by this work is described in chapter III. Once the concentration signals are detected as abnormal by the detection algorithm, the GPRS module will send out an audible alarm and warning messages to the maintenance person. It works in the 900MHz band. It will also send the collected data and the diagnostic result to the cloud server via Internet.

Wireless Sensor Nodes Design

All the data analyzed by the detection system is acquired by the sensor nodes. So the impact of the environment factors on the sensor data will be mapped to the result of the detection system. At this point, the temperature, humidity and pressure sensors (THPs) are included in the nodes. The sensor nodes in this work are designed by MCU, ZigBee wireless transceiver module, sensors module, and power supply module.

Considering the feasibility and safety of experimental operations, ethanol is used as the test gas and the sensor TGS2620 was selected, with a 50-5,000 part per million (ppm) measurement range and a 0.3-0.5 sensitivity (resistance ratio). The sensor is comprised of a metal oxide semiconductor layer formed on an alumina substrate together with an integrated heater. In the presence of a detectable gas, the sensor's conductivity increases as the gas concentration increases. A simple electrical circuit can convert the change in conductivity to an output signal which corresponds to the gas concentration. For the test, it consumes an average of 42mA current at 5V. The power consumption of gas sensor is approximately 210mW. To meet power requirement of sensor node, it would be possible to utilize a solar scavenger with a rechargeable battery, and combined with adjustable sampling frequency according to gas leak or not, and unnecessary alarms report to minimize the energy consumption. As efficient microcontrollers become cheaper and less power hungry, the only component left to be improved is the sensor [24]. The main weakness of metal oxide semiconductor gas sensor used in the paper is its large energy consumption, which is also common shortcoming of current gas sensors. Low-power gas sensors are currently less commercially available, but the gas sensors reported in the literature have been able to achieve very low power consumption. For example, Yeon Hoo Kim et al. [25] demonstrated self-activated transparent all-graphene gas sensor, the power consumption is 12μW to 14.2mW for the applied voltage of 1 to 60V. The power consumption of Low power wireless gas sensors used by Andrey Som et al. [26] is 75mW in the continuous measurement. Its low power consumption is achieved by applying a heater implemented as 10 μm platinum microwire in glass insulation. The literature [27] realized gas sensor by integrating the novel 3D hybrid aerogel on a low-power Micro-heater platform, and when the temperature of Micro-heater is 200^{o}C, the power consumption is only 4mW. In the future, the commercial application of these low-power gas sensors will greatly improve the continuous working time of wireless gas sensor nodes.

The scholars also researched the other strategies to solve the energy consumption and power supply problems of wireless sensor nodes. The literature [28] studied the system- and circuit-Level optimization of power supply system for Wireless Sensor Networks, and renewable power-supply system. Mingyi Chen [29] presented a self-powered wireless sensor node powered by electromagnetic energy harvester, the results show that designing chip architecture with less components is also good strategy.

The algorithm mentioned in the paper is suitable for low sampling rates to conserve the AD sampling's power consumption and when detecting a sudden gas leak, then increase sampling frequency to ensure the sensor data accuracy. To increase the continuous working time of wireless gas sensor node, choosing the large capacity lithium cells is also one compromise method. For example, when the gas sensors work intermittently, and it works 10s every 20 seconds for data acquisition when have no gas leakage. Based on the 10000mAh size lithium cells available on the market, the gas sensor can work for about 40 days. If the photovoltaic solar panel and future low power consumption gas sensors with a few μw to mw order of power consumption will be used in the sensor nodes, a longer working time can be reached.

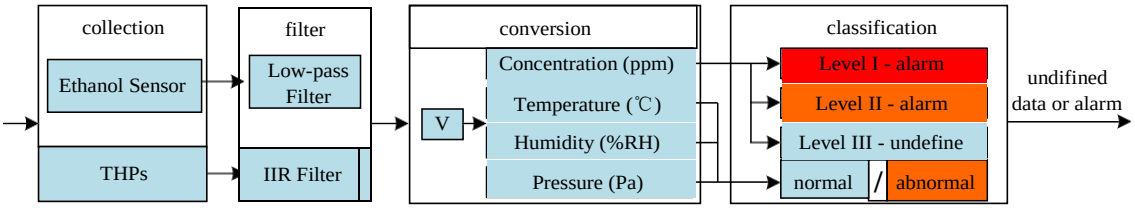


Fig. 2. The flow diagram of the data processing in the sensor nodes.

For the THPs, the BME280, which is a MEMS sensor integrated temperature, humidity and pressure, is adopted. The sensor has a built-in IIR filter to filter short-term disturbances. It consumes 11.88μW when it works and 0.33μW when it is dormant. For the MCU and wireless transceiver module, the highly integrated CC2530F256 chip is used as the node processor and transceiver. In order to achieve the minimum power consumption of the node, we reduce its transmission power under the condition of satisfying the normal monitoring of the system, where the transmission distance is 70 m. The peak of working current specified for CC2530F256 chip is 79 mA at 3.3 V supply in continuous measuring mode. The major function of this chip includes the collection of signals, filtering, Voltage/Indicators conversion, classification and the signal transmission. The flow diagram of the data processing is shown in Fig. 2.

In addition, the sensor node is powered by a lithium battery and the BP24210 solar charging module produced by TI. A photography of the sensor node can be seen in Fig. 3.

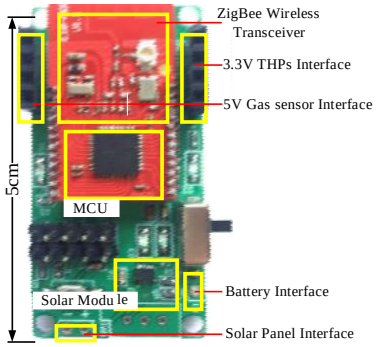


Fig. 3. The physical diagram of the sensor node.

The Deployment of Sensor Nodes

The placement of the sensor is critical to the detection result. It affects the collection of data directly, even if the sensor is located near the gas source. The sensor nodes can be distributed randomly, deterministically, or uniformly. On the premise of being familiar with the monitoring area, the deployment of sensors is deterministic. The sensor nodes should be deployed at the location where the concentration data is prone to be collected. The simulations for the leak of volatile ethanol by Fluent are performed to analyze the distribution of the concentration and further determine the position of the sensor under different conditions. Fig. 4(a) presents that the leaking gas is influenced by the wind direction, the source nozzle size is 6mm and the wind direction is set from left to right. It is observed that the gas diffusion is divided into four stages: first, the gas diffuses in the initial injection direction; second, the gas spreads mainly in the downwind direction under the influence of the wind; then the gas sinks and accumulates as the density

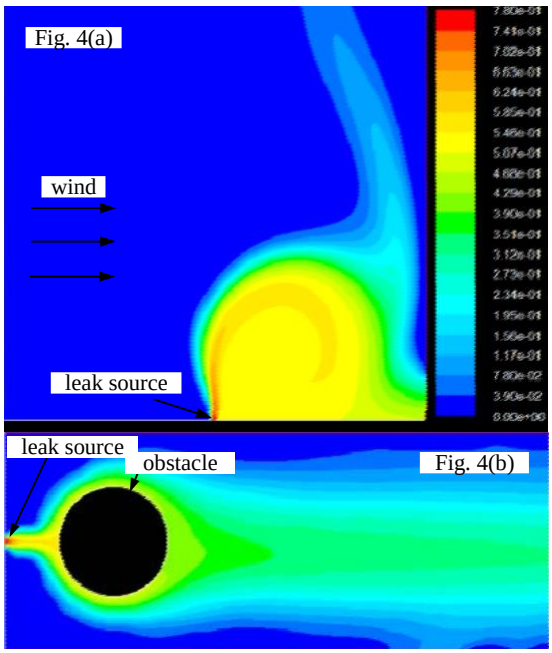


Fig. 4(a). The gas diffusion simulation under the influence of the wind direction and gas characteristics.
Fig. 4(b). The concentration distribution in presence of an obstacle near the source of the leak.

of ethanol is greater than the air's; third, the buoyancy force plays a leading role and the gas diffuse upward; finally, the gas concentration is negligible when the gas is diluted in air. And Fig. 4(b) illustrates the concentration distribution in presence of an obstacle near the leak source. The leaking gas bypasses the obstacle or accumulates nearby the obstacle. There has higher concentrations at both front and back of the obstacle.

From Fig. 4(a), the sensor is mainly collected during the second stage of gas diffusion, since the initial injection direction is unpredictable and the concentration is too low in the latter two stages.

Therefore, combined with the simulation results and the safety specification of combustible gas detection, the deployment rules of the sensor nodes are summarized as follows:

- The more sensor nodes should be installed at the downstream of the high frequency wind direction relative to the potential leak source;
- When the density of the gas is greater than 0.97kg/m^3, the sensor nodes should be deployed below the level of the potential leak source;
- Considering the presence of the obstacle, the sensor should be placed in the vicinity of the obstacle (e.g. columns or walls);
- A suitable obstacle will be placed near the sensor nodes to increase the gas residence time when the monitored area is open or the wind speed is high;

III. The Method of the Gas Leak Detection

A. Time-Domain Features Extraction

Time-domain features such as variance, mean, kurtosis or skewness are not suitable as a basis for judging between normal and abnormal. The main drawback of mean is that it is easy to be affected by extreme values. By increasing the time window of each average, the effect of extreme values can be reduced, but the monitoring delay of the system is increased. Variance can indicate the deviation of all samples from the mean value to a certain extent, but it can not be used as a parameter for rapid detection of gas leakage. Kurtosis are statistics describing the overall extent of the slow distribution of all forms steep values, which is not suitable as a statistical parameter for gas leakage. Skewness is a measure of the skewness and extent of statistical data distribution. It is a numerical feature of the degree of asymmetry of statistical data distribution, and is not suitable for dynamic feature statistics. Mean and variance is used in ACFs. They are part of the algorithm. These parameters, such as mean, variance, is not suitable for determining whether a gas is leaking or not. The algorithm is supposed to detect the leaking as soon as possible, but these parameters are time-consuming.

When no leaks occur, the concentration of the monitored gas approximates to zero. And if leaks happen, the amplitude of concentration will fluctuate and deviate from zero. Therefore, a common detection method of gas leak is to set fixed alarm concentration point (such as 25%LEL and 50%LEL). The problem is that the concentration is difficult to accumulate to the alarm points when the leak aperture is small. Thus, the non-concentration threshold need to be set.

The measurement curves of gas sensors in the process of gas leakage and non-leakage， as well as the sampled normal and abnormal signal of concentration in three minutes are shown in Fig. 5. According to product information of TGS 2620, this sensor is insensitive to low level ppm of ethanol. When the gas concentration is lower than 50ppm, the sensor has a lower precision because of the insensitivity. The sampling rate of the sensor is 1Hz during the experiment. There is no need for high sampling rate in the detection process of the system, which is helpful to reduce the power consumption. The abnormal signal was generated with a man-made ethanol gas leak, whose aperture is about 6mm. The ethanol sensor is placed downwind at a distance of 20cm from the leak source made by our laboratory.

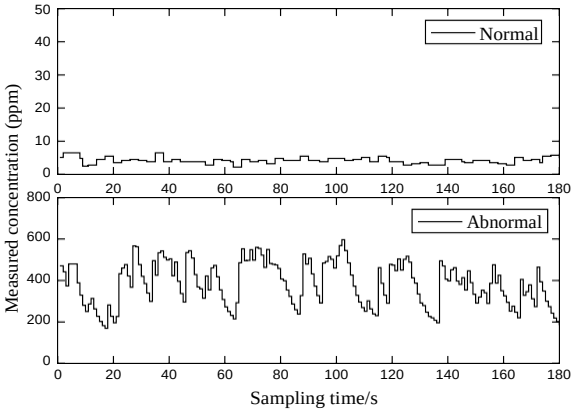


Fig. 5 Top: normal concentration data (without leak); bottom: abnormal concentration data (with leak).

The signals of normal and abnormal are divided into three segments respectively, and then the ACF of each segment is calculated. The result is shown in Fig. 6.

The similarity between ACFs is determined by the correlation coefficient, which is quantified by:

$$rr_{iiii} = \frac{CCCCCC(AA_{ii}, AA_{ii})}{\sigma\sigma_{ii} \cdot \sigma\sigma_{ii}} = \frac{EE((AA_{ii} - \mu\mu_{ii})(AA_{ii} - \mu\mu_{ii}))}{\sigma\sigma_{ii} \cdot \sigma\sigma_{ii}} \quad (1)$$

where AA_{ii} and AA_{ii} are two ACF, $\mu\mu_{ii}$ and $\mu\mu_{ii}$ are the means of AA_{ii} and AA_{ii}, $\sigma\sigma_{ii}$ and $\sigma\sigma_{ii}$ are the standard deviations of AA_{ii} and AA_{ii}. In this way, the correlation coefficients rr_{iiii} are normalized to the range of 0 to 1 firstly and then a correlation matrix composed of rr_{iiii} is formed. Each segment is characterized by a degree value, which is obtained by calculating the mean of the coefficients for each column of the matrix. When the segments are all normal, the degrees of them are high. And when the segments contain many normal segments and an abnormal segment, the degree of the abnormal segment is low. The degree feature is independent of the absolute amplitude and reflects the similarity between an undefined segment and normal ones. Therefore,

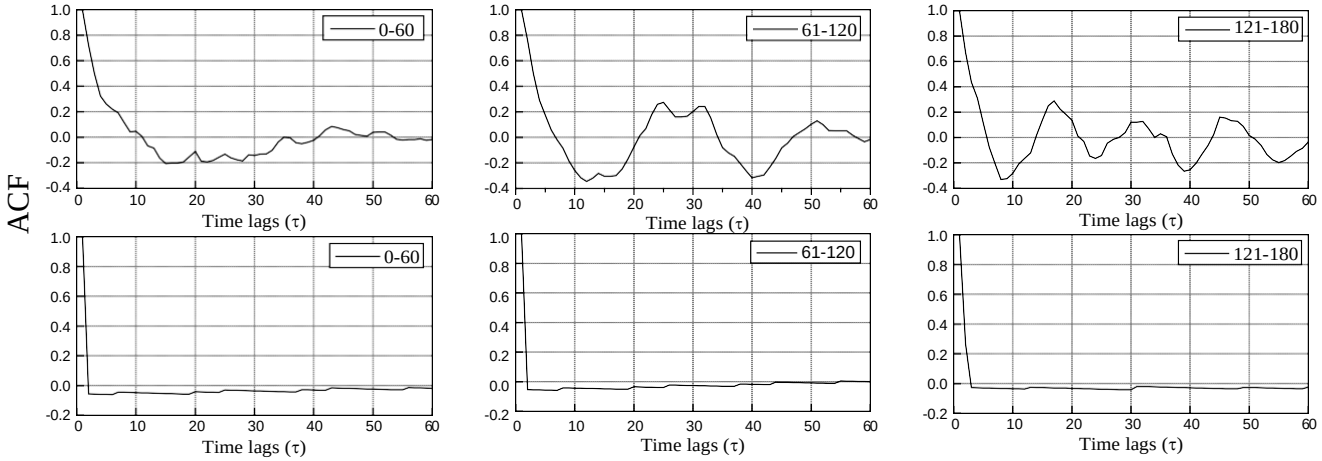


Fig. 6. ACFs of six different concentration segments of on minute. When the signal is normal, their ACFs are similar and stable while the abnormal ACFs have large fluctuations which are different strongly from the normal ACFs. This can be seen in Fig. 10 where the ACFs of three normal segments (bottom) and three abnormal segments (top) are plotted together.

the degree can be used as the feature to differentiate the normal and abnormal signals.

B. Implementing Leak Detection with Time-Domain Features

1) The General Procedure: For the system implementation, the leak detection model is established with the normal concentration data and applied to the real-time gas leak detection. Fig. 7 illustrates the leak detection procedure used by this study.

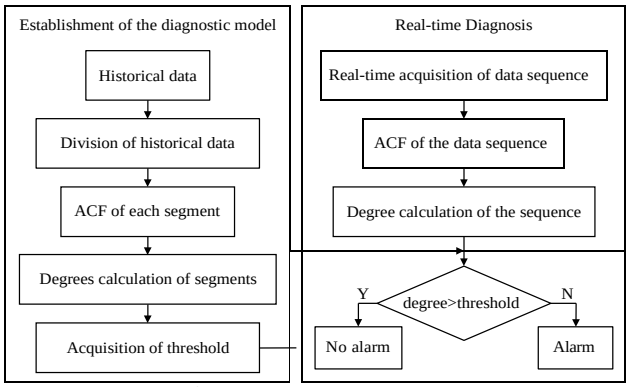


Fig. 7. The procedure for real-time leak detection

2) Feature Extraction: The training historical data took two hours to complete the sampling continuously when no leak. The time-domain features, which described in section III-A, can be extracted by the following steps.

a. Divide the historical data into *M* segments with a step value *N* (with *N*=30).
b. Then the ACF of each segments is calculated.
c. Next, the correlation matrix is formed according to the equation (1).
d. Finally, the degrees RR_{ii} of ACFs are computed by averaging each column of the correlation matrix:

$$RR_{ii} = \frac{1}{MM-1}\sum_{ii=1}^{MM-1} (ii \neq jj) \qquad (2)$$

3) Threshold Determination: In the previous step, nearly three hundred degrees have been obtained. These degrees are used to do the demonstration of the method and more historical data is needed in practice. The histogram of the degrees' distribution can obtained by the following steps:

a. Divide the value of degrees into MM_0 groups with a step value $AA_{sssssss}$, which can be written as:

$$MM_0 = \frac{KK_{mmmmmm} - KK_{mmiimm}}{AA_{sssssss}} \qquad (3)$$

Where RR_{mmmmmm} and RR_{mmiimm} are the maximum and minimum of the degrees, respectively. The value changes in the second place after the decimal point, so the step $AA_{sssssss}$ is set to 0.05;

b. Count the number of degrees whose value fall in the range of the kkth ($k = 1, \cdots, MM_0$) group and scale it as frequency;
c. Calculate the fitting function of the histogram distribution based on the statistics of the degrees;

Plot the statistical distribution curve of the degrees and the fitting function;

Through the process of the fitting, the probability density function (PDF) of the degrees is obtained as is shown in Fig. 8.

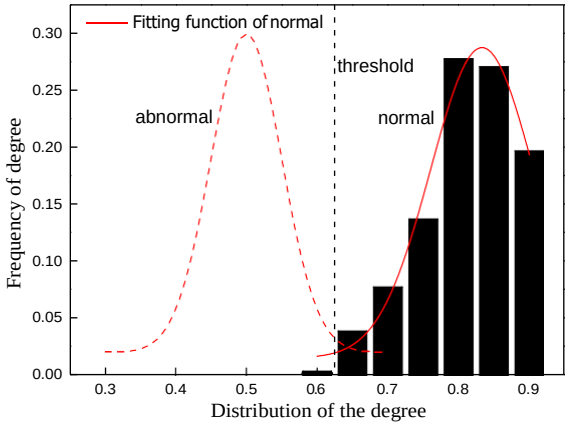


Fig. 8. The statistical distribution curve of degrees.

In other words, the probability $PP(RR_{ii}|\theta\theta = 0)$ against the degree RR_{ii} is obtained when no leak ($\theta\theta = 0$). According the previous theories described in section III-A, the degrees of the abnormal segment are less than the degrees of normal, generally, as is shown in the Fig. 9.

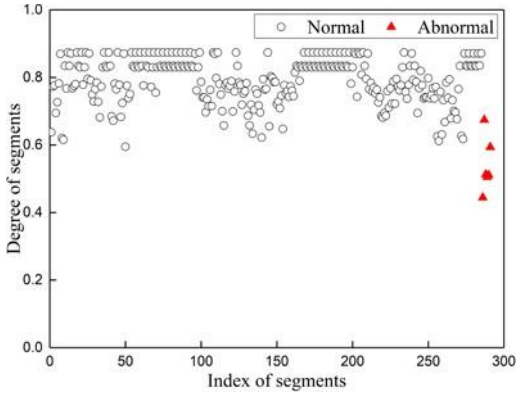


Fig. 9. The circle symbols represent normal degrees, while the triangle symbols represent abnormal degrees.

If enough leak data is available, the probability $PP(RR_{ii}|\theta\theta = 1)$ against the degree RR_{ii} also can be obtained when leak which is shown by the dashed line in Fig. 8. The sequential probability ratio (SPR) can be used to detect the leak [30]. However, it is almost impossible to get actual leak samples. So the threshold should be set so that the probability $PP(RR_{ii}|\theta\theta = 0)$ is low under the premise of higher detection rate and lower false positive. When the degree is the threshold, the probability $PP(RR_{ii}|\theta\theta = 0) = P_{sh}$.

The threshold is a critical parameter. The false negative judgment increases when the threshold is too low, while the false positive one rises when the threshold is too high. So a suitable balance between them needs to be confirmed.

4) Selection of Sensor Nodes and Fusion of the Real-time Data: Generally, the WSN uses a large number of sensors to obtain data. As the deployment grows in size, a faraway sensors from the leak source will probably not be able to detect any change. The response of the sensors to the plume of gas will depend on its location with respect to this plume, to confirm it, a test is completed by placing six sensors at different distances from the source of the leak in the downwind direction. It is found that the concentration decreases and phase of the detected concentration curve delay away from the source of the

leak, as shown in Fig. 10.

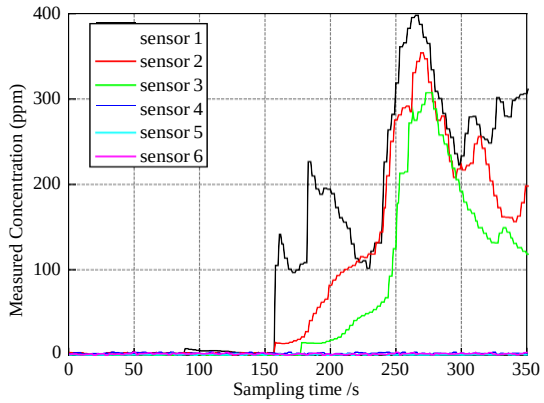


Fig. 10. The concentration curve of sensors at different distances from the leak source.

The several sensors closest to the leak source should make up a set to participate in the detection of gas leak such as sensor 1, 2 and 3. However, the location of the source leak is uncertain. The general rule is that the closer the distance to the leak source, the greater the gas concentration. So the sensor with the largest mean concentration over a period of time was assumed to be the virtual leak source (VLS).

In this work, the concentration data collected by the iith sensor at time tt is expressed as $xx_{ii}(tt)$ where i $\in \{1,2,\ldots,\mathrm{n}\}$, n is the total number of sensors that detected the third level data. The distance $dd_{iiiiiii}(tt)$ between the VLS sensor and sensor i can be written as:

$$dd_{iiiiiii}(tt) = \exp\left[-\left(xx_{ii}(tt) - xx_{iiiiii}(tt)\right)^2\right] \quad (4)$$

The smaller $dd_{iiiiiii}(tt)$ is, the greater the distance is. In addition, the distance $dd_{iiiiiii}(tt)$ is also fluctuant since the concentration is fluctuant. At a moment, a sensor is close to the VLS sensor, and it may be far away from the VLS sensor at the next moment. So when the mean of $dd_{iiiiiii}(tt)$ is large and its variance is small, the sensor i is considered close to the leak source in the period of time T. The distance $DD_{iiiiiii}$ is given as:

$$DD_{iiiiiii} = [1 - 0.05\sigma\sigma_{ii}^2] \cdot \overline{dd}_{iiiiiii} \quad (5)$$

And there are

$$\overline{dd}_{iiiiiii} = \frac{1}{NN}\left[dd_{iiiiiii}(tt_1) + dd_{iiiiiii}(tt_2) + \cdots + dd_{iiiiiii}(tt_{NN})\right] \quad (6)$$

$$\sigma\sigma_{ii}^2 = \frac{1}{NN}\sum_{kk=1}^{NN}\left[dd_{iiiiiii}(tt_{kk}) - \overline{dd}_{iiiiiii}\right]^2 \quad kk \in (1, NN) \quad (7)$$

Where N is the step value and the number of sampling points in the period of time T.

Thus, a set of distance is given as:

$$DD_{mm} = \{DD_{1iiiiii}, DD_{2iiiiii}, \cdots, DD_{mmiiiiii}\} \quad (8)$$

According to the value of $DD_{iiiiiii}$, the top m sensors are selected into the new set DD_{mm} that is given as:

$$DD_{mm} = \{DD_{1iiiiii}, DD_{2iiiiii}, \cdots, DD_{mmiiiiii}\} \quad (9)$$

Correspondingly, the set SS_{mm} including s sensors is obtained as:

$$SS_{mm} = \{ss_1, ss_2, \cdots, ss_{mm}\} \quad (10)$$

The weighting coefficient cc_{ii} of each sensor in the set SS_{mm} is calculated according the $DD_{iiiiiii}$ in the set of DD_{mm} as:

$$cc_{ii} = \frac{DD_{iiiiiii}}{\sum_{ii=1}^{mm} DD_{iiiiiii}} \quad (11)$$

Hence the data from the sensors close to the VLS play a more important role in the leak judgment. Finally, the value $\mathrm{X}(tt)$ in the CS, the result of fusion, is given as:

$$\mathrm{X}(tt) = \sum_{ii=1}^{mm} cc_{ii}\, xx_{ii}(tt) \quad (12)$$

The CS is recognized as the real-time data sequence and is used as the input of the diagnostic model.

5) Leak detection: Once the real-time CS is obtained, its ACF is computed firstly. According to the flow in Fig. 11, the real-time degree needs to be obtained. At the same time, several ACFs of the normal sequence are required for the calculation of the real-time degree. The length of data in one detection, denoted as L, is divided into MM_1 segments (named window 1). As a result, the length of each segment is $NN = LL/MM_1$ (named window 2) which is the same as the value mentioned in the step 2. The first MM_1-1 ACFs are taken as the reference values and the M_1th is the real-time ACF. So its degree is obtained according to the equation (1) and (2). If the new degree is less than the threshold, it means the input concentration is abnormal; otherwise, the concentration is normal.

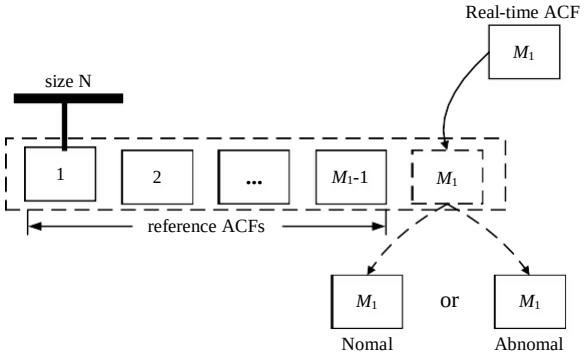


Fig. 11. The diagram of the real-time detection.

IV. Experimental Validation

A. *Verification of the effect of ambient temperature on the output response of the sensor*

As MOX sensor, the output of the gas sensor also depends on the temperature of the internal heater, we have done some experiments on the gas to analyze the effect of ambient temperature on the output response of the sensor. We have measured internal heater temperature and the response of the gas sensor at multiple temperatures in the temperature box. Experimental devices are shown in Fig. 12. Experimental device structure diagram is shown in Fig. 13.

The effect of working environment temperature on MOX gas sensor is shown in Fig. 14 with a red line. When the gas sensor is placed in the thermostat, the temperature of the gas sensor heater rises as the temperature of the thermostat rises, when the ambient temperature reaches 47.5℃, the heater temperature rises to 82.5℃. The output voltage of the sensor corresponding to the test temperature is shown in Fig. 14 with a blue line. The experimental results show that the output variance of the sensor is 0.126 in the working environment temperature range of 25℃ to 47.5℃.

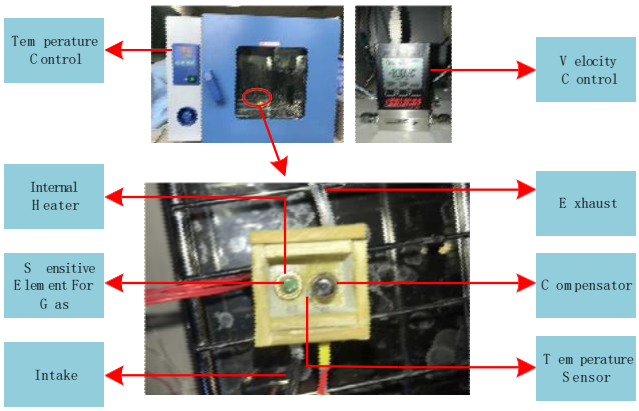


Fig. 12. Experimental devices

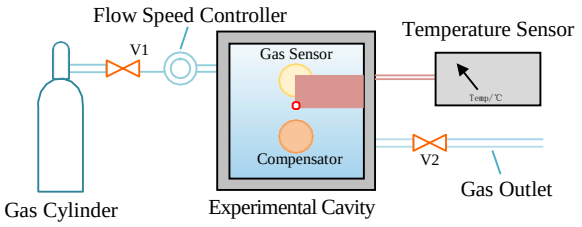


Fig.13 Experimental device structure diagram

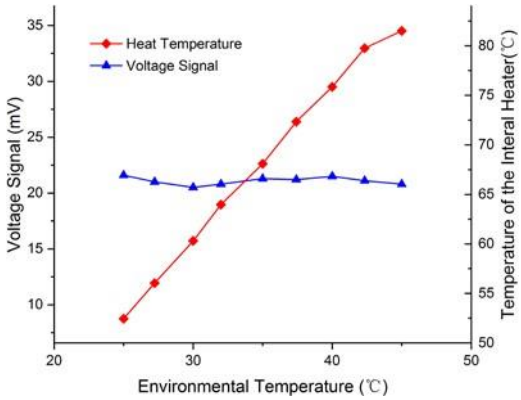


Fig. 14 Effect of ambient temperature on temperature of internal heater

It can be seen from the diagram that the heater temperature in gas sensor increases with the environment temperature, but the effect of environment temperature on the gas sensor is relatively small. This is because of the existence of compensator inside the sensor, the external temperature has little interference on the sensor. As we can see from Fig. 12, the compensator that is wrapped in an aluminum shell has the same physical structure as the detector. So the only difference between them is the composition of the gas around the two detectors. The influence of external factors on test results could be reduced by such structure.

According to working mechanism of MOX gas sensor [27], heating the sensing material of MOX sensor with a heater can enhance the reversibility of the sensor and accelerate the response and recovery rates. The response and recovery time might be faster at higher temperature than room temperature, therefore, the performance of MOX gas sensor is affected by the heater temperature in gas sensor. But according to the tested results, the working environment temperature changes will has little effect on sensor performance for MOX gas sensors with one good temperature compensating structure.

The effect of humidity on the performance of MOX gas sensor has not been tested due to the lack of testing setup. According to literature [31], humidity effects maybe neglected as they are usually an order less significant than temperature effects which is due to temperature change in the sensor heating element

B. Field Tests of Leak Detection System

In order to verify our system, the experiment was carried out, as shown in Fig. 15. For safety, the volatile and non-toxic alcohol (95% alcohol purity) is selected as the source of gas. The source of the leak consists of air pump, power supply, alcohol bottle and nozzle. In the leak process, the gas concentration is constant. Hence, the leakage only depends on the size of the nozzle. In the experiments, the nozzle sizes are 6 mm, 11 mm, 20 mm. 30 leaks were created and each one lasted 3 minutes. It takes about five minutes to initialize the gas sensor before testing (sensor resistance preheating), and the node will work well after a preliminary classification and Zigbee restart needs time until become active, it's all about 10 seconds to start measuring in the system.

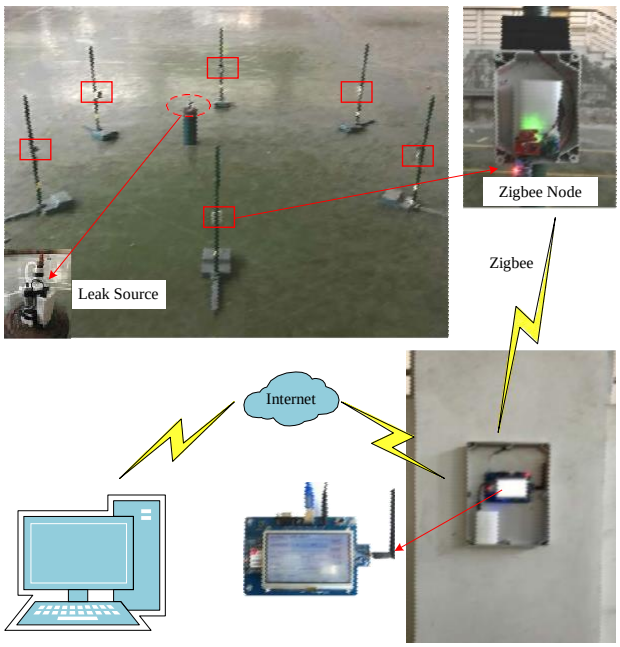
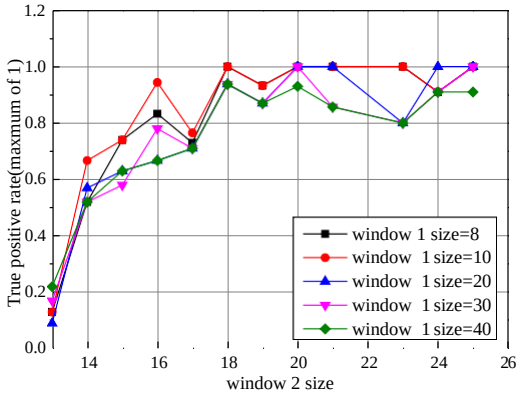


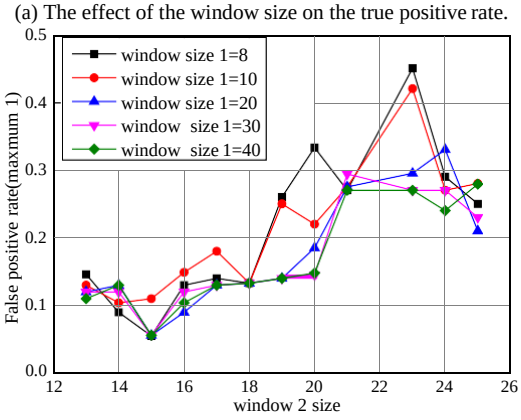
Fig. 15. Site of the experiment

C. Parameters Selection for Gas Leak Detection

Several parameters have significant effects on the judgment in the proposed method. They are analyzed respectively. As shown in Fig. 16 (a), the general trend observed show that the true positive rate of the detections increases as the length of the window 2 sizes increases greatly. And the impact of the size of window 2 on the true positive rate tends to be stable beyond a particular point (around 18). The length of window 1 shows little influence on the true positive rate of the detections. The impact of changing the window 1 size and window 2 size on the false positive rate is shown in Fig. 16 (b). It demonstrates that the false positive rate fluctuates at a lower level when the size of window 2 is small and then increases significantly when the length of window 2 exceeds a value (about 18). The trend shows that increasing the window 1 size does not affect the false positive rate obviously, except when it reduces to a small value (such as 8 or 10).



(a) The effect of the window size on the true positive rate.



(b) The effect of the window size on the false positive rate.

Fig. 16. The effect of the window size on both the true and false positive rate.

Besides the window sizes, the threshold is an important parameter that determines the performance of the detection method. The effect of the threshold on both the true and false positive rate can be seen in Fig. 17.

According the general trend, the true and false positive rate increase drastically. Furthermore, there are no detectable leaks and false positive judgment when the threshold is particularly small (set to 0). In addition, almost all of the segments will be misjudged (false positive rate of 100%) in another extreme case in which the threshold set to 1. The true positive rate is close to 100% at the same time because any degrees of the leak segments are certainly smaller than 1. However, false positive rate of 100% is unacceptable. Therefore, the threshold should be set to decrease the false positive rate on the premise of ensuring high true positive rate. In Fig. 17, the threshold is set at around 0.68 and the corresponding probability P_{sh} is 0.02.

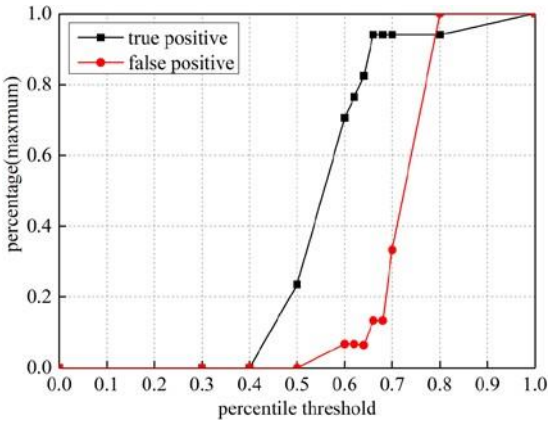


Fig. 17. The effect of the threshold on both the true and false positive rate.

D. Results

According to the parameters selection rules presented in subsection IV-C, the parameters are set as $NN = 18, MM_1 = 40$. In this test, historical data were collected for one day offline and more than four thousand segments were got as training samples when no leaks. Then the statistical distribution curve of degrees. After fitting, the PDF curve $PP(RR_{ii}|\theta\theta = 0)$ is obtained. So the threshold is set as 0.68 when the probability P_{sh} is 0.02. By using real-time data processing methods, real-time degrees are obtained and the results of leak detection are shown in the column of Table I. Based on the field operation and corresponding inspection records, during the monitoring of the set of 50 sets of leakage test group and 50 sets of non-leakage test group, the method of this paper successfully detected 47 sets of leakage data. When detecting the non-leakage data, 5 sets of false alarms occurred, and the delay was within 30S. Furthermore, the method of likelihood function [24] is used for comparison. A pipeline small leak detection by harmonic wavelet is used in the contrast experiment. The experimental setup of comparative approach is as follow:

- Stage 1 window size: 20;
- Stage 2 window size: 150;
- Percentile threshold: 15%.

The PDF in comparative approach is obtained based on the statistics distribution of the historical data in preparation stage. And the likelihood of sensor data is obtained. By dividing the segments, the similarity of the segments is calculated. The percentile threshold are determined by observing the general trend and keeping trying. The detection result are given in the Table I. There are 42 true positive alarms and 7 false alarms. It is obvious that the proposed method in this work is better than the comparative method in the term of detection efficiency. Moreover, the quite good performance is obtained in the detection delay. The time of the proposed approach is within 30s while the comparative approach is more than 100s averagely. In general, the method proposed by this study is more efficient.

TABLE I
THE COMPARISON OF TEST RESULTS BETWEEN COMPARATIVE AND PROPOSED METHODS

Signal type	Comparative method	Proposed method
Artificially leak signal	50	50
True positive alarm	42	47
Experimental group of no leakage	50	50
False positive alarm	7	5
Delay time	≥100	≤30

V. CONCLUSION

In this paper, a concentration-based leak detection method with statistical feature extraction is proposed. The detection model is built with the feature extracted from normal concentration signals. The approach obtained the non-concentration threshold using the distribution of the correlation between ACFs of normal concentration segments. Moreover, an automatic procedure is described to select sensor nodes that participate in leak detection considering the location of sensors. And the real-time concentration is the last segment in the real-time diagnosis process, which reduced greatly the detec- tion delay and calculation. Additionally, the analysis of the position of the sensors by the simulations of Fluent is performed in consideration of the influence of wind direction and obstacles during the deployment of the sensor network. Modeling simulation results provide recommendations for node installation. And the sensors are placed at the points where the concentration of gas is more easily collected. Based on the field operation (with 8 sensors and a monitoring area of 50 mm^2) and corresponding inspection records (with an average delay of 30 s), there were 47 true positive alarms,5 false positive alarms with the proposed method. The experimental results demonstrate that the proposed method in this work can effectively detect the gas leaks.

In the future, many topics remained for research and gas leak detection will make great progress. The noise in the original signals has great impact on the detection of small leaks and should be considered. With the improvement of wireless communication reliability, and as efficient microcontrollers become cheaper and less power consumption, the only component need to be improved is the sensor. The power consumption of gas sensors is a huge obstacle to the application of WSN in the field of gas leak detection and a gas sensors can only measure one gas; Detection of multiple gas types will result in increased energy consumption of nodes, which provides an opportunity for development of the energy harvesting and MEMS process. Better sensor hardware can provide more accurate data for algorithms.

REFERENCES

[1] State Administration of Work Safety, “Statistics of hazardous chemical accidents”, State Administration of Work Safety, retrieved on Dec. 2017[Online]. Available: http://www.chinasafety.gov.cn/newpage/wxhxp/wxhxp2013_jdys.htm

[2] Murvay P S, Silea I, “A survey on gas leak detection and localization techniques,” Journal of Loss Prevention in the Process Industries, vol. 25, no. 6, pp. 966-973, 2012.

[3] S P. Spachos, L. Song and D. Hatzinakos, "Gas leak detection and localization system," *2014 IEEE 11th Consumer Communications and Networking Conference (CCNC)*, Las Vegas, NV, 2014, pp. 1130-1131.

[4] S L. Shu, M. Mukherjee, X. Xu, K. Wang and X. Wu, "A Survey on Gas Leakage Source Detection and Boundary Tracking with Wireless Sensor Networks," in *IEEE Access*, vol. 4, pp. 1700-1715, 2016.

[5] W H. Wang, L. Dong, W. Wei, W. S. Zhao, K. Xu and G. Wang, "The WSN Monitoring System for Large Outdoor Advertising Boards Based on ZigBee and MEMS Sensor," in *IEEE Sensors Journal*, vol. 18, no. 3, pp. 1314-1323, Feb.1, 1 2018.

[6] A. Somov, A. Baranov, D. Spirjaki, A. Spirjakin, “Deployment and evaluation of a wireless sensor network for methane leak detection” *Sensors & Actuators a Physical*, vol. 202, no. 10, pp. 217-225, 2013.

[7] V. Jelicic, M. Magno, D. Brunelli, G. Paci and L. Benini, "Con-text-Adaptive Multimodal Wireless Sensor Network for En- er-gy-Efficient Gas Monitoring," in IEEE Sensors Journal, vol. 13, no. 1, pp. 328-338, Jan. 2013.

[8] D. Brunelli, M. Rossi, "Enhancing lifetime of WSN for natural gas leakages detection", Microelectron. J., vol. 45, no. 12, pp. 1665-1670, 2014.

[9] Aliyu F, Al-Shaboti M, Garba Y, et al, “Hydrogen Sulfide (H2S) Gas Safety System for Oil Drilling Sites using Wireless Sensor Network” *Procedia Computer Science*, vol. 63, pp. 499-504, 2015.

[10] Zhao L, Yang H, “Small-target leak detection for a closed vessel via infrared image sequences,” *Infrared Physics & Technology*, vol. 81, pp. 109-116, 2017.

[11] Universidad Carlos III de Madrid, “SENSIA’s Gas Imaging System,” Universidad Carlos III de Madrid, 2015.

[12] J. Leis and D. Buttsworth, "A Robust Method for Tuning Photoacoustic Gas Detectors," in *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4338-4346, May 2018.

[13] J. Leis and D. Buttsworth, "A Temperature Compensation Technique for Near-Infrared Methane Gas Threshold Detection," in *IEEE Transactions on Industrial Electronics*, vol. 63, no. 3, pp. 1813-1821, March 2016.

[14] Emerson, “GDU-Incus Ultrasonic Gas Leak Detector Emerson Process Management,” Emerson, 2014.

[15] Akouemo H N, Povinelli R J, “Probabilistic anomaly detection in natural gas time series data,” *International Journal of Forecasting*, vol. 32, no. 3, pp. 948-956, 2016.

[16] P. Gupta, A. Goyal, J. Dauwels and A. Ukil, "Bayesian detection of leaks in gas distribution networks," *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*, Florence, 2016, pp. 855-860.

[17] F. Wang, W. Lin, Z. Liu, S. Wu and X. Qiu, "Pipeline Leak Detection by Using Time-Domain Statistical Features," in *IEEE Sensors Journal*, vol. 17, no. 19, pp. 6431-6442, Oct.1, 1 2017.

[18] Xiao Q, Li J, Bai Z, et al, “A Small Leak Detection Method Based on VMD Adaptive De-Noising and Ambiguity Correlation Classification Intended for Natural Gas Pipelines,” *Sensors*, vol. 16, no. 12, pp. 2116, 2016.

[19] Y. Yingchun and Z. Laibin, "De-noising of acoustic signal by natural gas pipeline leakage based on DTCWT-SVD," *2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA)*, Hefei, 2016, pp. 1551-1556.

[20] Hu J, Zhang L, Wei L, “Detection of small leakage from long transportation pipeline with complex noise,” *Journal of Loss Prevention in the Process Industries*, vol. 24, no. 4, pp. 449-457, 2011.

[21] J. Kang, Y. J. Park, J. Lee, S. H. Wang and D. S. Eom, "Novel Leakage Detection by Ensemble CNN-SVM and Graph-Based Localization in Water Distribution Systems," in *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4279-4289, May 2018.

[22] C. Varon, D. Testelmans, B. Buyse, J. A. K. Suykens and S. Van Huffel, "Robust artefact detection in long-term ECG recordings based on auto-correlation function similarity and percentile analysis," *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Diego, CA, 2012, pp. 3151-3154.

[23] M. Lavielle, "Optimal segmentation of random processes," in *IEEE Transactions on Signal Processing*, vol. 46, no. 5, pp. 1365-1373, May 1998.

[24] F. Chraim, Y. Bugra Erol and K. Pister, "Wireless Gas Leak Detection and Localization," IEEE Transactions on Industrial Informatics, vol. 12, no. 2, pp. 7*68-779, April 2016*.

[25] Kim, Y.H.; Kim, S.J.; Kim, Y.-J.; Shim, Y.-S.; Kim, S.Y.; Hong, B.H.; Jang, H.W. Self-activated transparent all-graphene gas sensor with endurance to humidity and mechanical bending. ACS Nano 2015, 9, 10453–10460.

[26] A. Somov, E.F. Karpov, E. Karpova, A. Suchkov, S. Mironov, A. Karelin, A. Baranov, D. Spirjakin, Compact low power wireless gas sensor node with thermo compensation for ubiquitous deployment, IEEE Transactions on Industrial Informatics, 2015,vol.11, no.6, pp.1660-1670.

[27] Hu Long, Anna Harley-Trochimczyk, Thang Pham, Zirong Tang, Tielin Shi, Alex Zettl, Carlo Carraro, Marcus A. Worsley, and Roya Maboudian, “High Surface Area MoS2/Graphene Hybrid Aerogel for Ultrasensitive NO2 Detection”, Advanced Functional Materials, 2016, pp.1-6.

[28] Ioannis Mandourarakis, Eftichios Koutroulis, Unified System- and Circuit-Level Optimization of RES-Based Power-Supply Systems for the Nodes of Wireless Sensor Networks, IEEE Transactions on Industrial Informatics, VOL. 14, NO. 2, FEBRUARY 2018, p598-607.

[29] Mingyi Chen, Min Wang, Hengwei Yu, Guanghui He, Yongxin Zhu, Yadong Liu, and Guoxing Wang, A Self-Powered 3.26- μ W 70-m Wireless Temperature Sensor Node for Power Grid Monitoring, IEEE Transactions on Industrial Electronics, Vol. 65, No. 11, November 2018, pp.8956-8965.

[30] F. Chraim and K. Pister, "Wireless valve position monitoring: A MEMS approach," *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society*, Vienna, 2013, pp. 4016-4021.

[31] N. Masson, R. Piedrahita, M. Hannigan, Approach for quantification of metal oxide type semiconductor gas sensors used for ambient air quality monitoring, Sensors and Actuators B, 208 (2015), pp.339 – 345

International Conference on Latest Trends in Electronics and Communication (ICLTEC) ISBN 978-93-88808-63-0

Influence Of Wind Power Plants On Power Systems Operation

Poushya Kandukuri 1 ***DRK Mahesh 2***

poushya99@gmail.com drkmahesh515@gmail.com

1,2 Assistant Professor , Malla Reddy College of engineering, Hyderabad,Telangana,India

Abstract— **In recent years, renewable energy sources (RES) has increased all over the world, especially in Europe, and the changes brought by these sources have a significant impact on system performance and efficiency. Integrating RES to an electric power network offers many techno-economical benefits and in the same time necessitate advances in the operation, control and planning of electric networks. This paper focuses on the influence of the interconnecting of wind power plants on the electric network in terms of optimal power system operation. In this paper is proposed a k-means clustering based approach to identify the typical load profiles for wind farms in order to achieve the optimal power system operation on minimizing the power losses.**

Keywords-wind power plant, optimal operation, typical load profiles, clustering techniques.

I. INTRODUCTION

The electricity industry restructuring, along with advances in small scale generation technologies, and a higher awareness of environmental issues are the key factors that have influenced the development of RES in the last period of time [1]. The electrical network of the future should be flexible, efficient, reliable and security of supply and in the same time will allow two way flows of energy and real time communication capable of self healing and enabling fast restoration from supply disruption facilitate market operations and customer choice and information [2].

The growing integration of the renewable energy sources like solar power, wind power, or combined heat and power in the power network impacts a lot of interested parties: transmission and distribution companies, the owners and operators of the distributed generation sources, other end users of the power network, regulators and policy makers.

An increasing introduction of RES without any changes in the electrical networks will ensue in unacceptable levels of quality and reliability. Small distributed generation sources (DG) are located to the medium or low voltage distribution network, where traditionally only consumption has been connected. The penetration of large quantities of them will require improvements not only at the voltage level where the sources are located but also at higher voltage levels. The intermittent character of renewable energy production introduces new power quality phenomena, representatively at lower voltage levels [3]. Besides, this variation as well as the difficulty in estimating RES production impacts the operation of the power transmission networks. These are the principal reasons why the RES integration is a very attractive research subject and it will continue being of great interest.

The changes brought by RES have definitely an important impact on system efficiency and performance and necessitate enhancements in operation and planning of electrical system. In order to maximize the potential RES benefits in the future could be taken enforcement measures or upgrading actions for improving the system performance and reliability. The upgrading process will be done, taking into consideration the electricity delivery infrastructure that is divided into transmission and distribution system, [4].

The traditional planning process for expanding transmission system is based mainly on its needs on past and designed loading levels, which have traditionally been assessed of future demand. In the deregulated market, and in the case of using different renewable energy sources, transmission planners must respond to the necessities of power generators. Otherwise, planning to develop transmission system may now be established by the location and type of generation source, rather than by the necessities of the transmission system.

Around the world, wind and photovoltaic power are considered to have the largest potential in electricity production from renewable energy sources, [5].

Currently wind energy has the largest interest because Romania has the highest potential from the Southeast of Europe, and Southeast of Romania ranks on the second place across the entire continent. In the past 10 years more studies considered necessary and appropriate addressing some activities of reassessment of wind potential of Romania, through the use of appropriate tools and instruments (measuring equipment, appropriate software) from measured wind data stations belonging to the National Agency of Meteorology, [6]. In [7] is given an estimate of the theoretical potential of wind energy corresponding (approximately 14 GW of installed capacity). Annual energy of this potential is 0,023 TWh / year.

Because in the last years there is an increase of wind farms connected to the grid powers, the electrical companies need by studies and analysis for to evaluate the impact of these sources on the system. These activities are performed in order to ensure reliable operation of the system in presence of wind farms [7], [8].

There are many requirements, prescriptions and guidelines that to refer to technical information necessary for assessment

of the influence on power system given by the wind sources. In the last years, there is a new approach that is based on the connecting to power systems of wind power farms with a large installed capacity (> 100 MW). Fig. 1 presents the influences of wind sources connected to the power systems, in function by the time horizon and dimension of surfaces that can be considered in the technical studies.

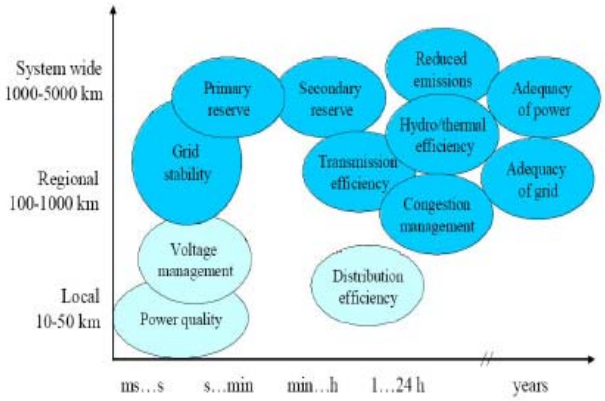


Figure 1. Influences of the wind sources connected to the power systems, [8]

The operation system of the wind turbine is conditioned by two parameters (speed wind - v and wind variations). There are three operating states that can be differentiated, [9], [10]:

Standstill of the turbine – for $v < v_{cut\text{-}in}$ or $v > v_{cut\text{-}out}$, where $v_{cut\text{-}in}$ and $v_{cut\text{-}out}$ represent cut-in and cut-out wind speeds .

Partial load – for $v_{cut\text{-}in} \leq v \leq v_n$, where v_n is the rated wind speed.

Full load – for $v_n < v \leq v_{cut\text{-}out}$. The value $v_{cut\text{-}out}$ is usually 25 m/s..

These three states represent characteristics P- v of the wind power plant. In order to study the influences of a wind sources on power system, it is needed to determine the patterns as accurately as possible. In the literature different techniques have been used for the classification and load profiling, but most of them were implemented to solve the problems from power systems. A review of the literature revealed two types of methods: statistical methods [11], [12] and methods based on artificial intelligence techniques fuzzy logic [13], [14], neural networks [15], data mining [16], clustering, [17]-[19].

In the paper an approach able to identify the optimal power systems operation considering the typical load profiles for wind farms is presented. First of all, an algorithm based on k-means method was used to find the typical load profiles for wind farms using a wind power generated database. After that, with this approach it can be find the optimal power system operation in order to minimize the power losses.

II. Determination Of Typical Load Profiles For Wind Power Plants

An approach based on clustering to determinate the wind power profiles for wind power farms from an electrical system is proposed. The *K*-means clustering algorithm used is used to classify operational profiles of wind power farms into coherent groups. By knowing these profiles, the operators can streamline the assessment of the demand.

The load profiling represents a different approach than the one based on metered demand. In this manner, for wind power farms from the electrical system are assigned a typical load profile. The shape of operational characteristic is influenced by the day (working or weekend) or season (spring, summer, autumn or winter). The operational characteristics of wind sources are in a very large number. This aspect create problems in analysis them. Thus, for an easy handling, they can be grouped in patterns, in function by the similarities between these. Every pattern will be characterized by a characteristic profile named typical load profile (TLP), [20], [21].

Each TLP is represented by a vector $x_i = \{x_{ih}, h = 1,\ldots, T\}$ for $i = 1, \ldots, K$, and the comprehensive set of TLPs is contained in the set $P = \{x_i, i = 1, \ldots, K\}$. The time scale along the day is partitioned into T time intervals of duration Δt_h , for $h = 1,\ldots, T$. Hourly values are used in this paper to exemplify the application. The variables used in the calculations are assumed to be represented as constant (average) values within each time interval. The clustering process forms K patterns corresponding the wind power plants. Further, the typical load profiles are assigned to wind power farms.

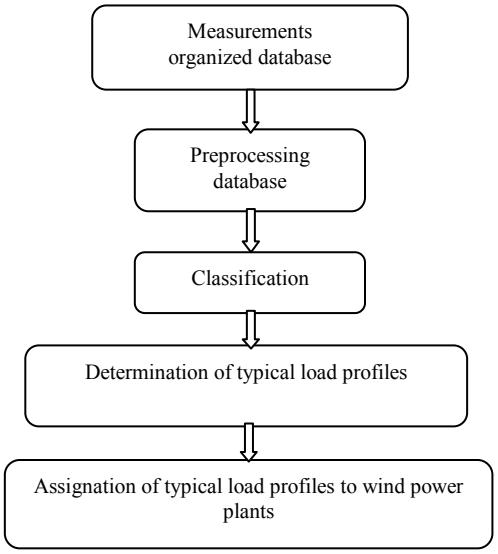


Figure 2. Flow-chart of the TPLs determination.

The proposed algorithm has the following steps, Fig. 2:

Step 1. Measurements: In this stage a database of operational characteristics is built.

Step 2. Data cleaning and pre-processing: A lot of technical aspects refer to communication problems, failure of equipment, etc., that can influenced negative the analysis need to be cleaned, pre-processed and reduced before the operational characteristics to be used in the clustering process.

Step 3. Classification: To realize this grouping, the K-means algorithm is used. Each operational characteristic is normalized. The normalization is made using the following relation:

$$z_h^{(j)} = \frac{x_h^{(j)}}{X^{(j)}} \quad ,j = 1, \ldots, N,\ h = 1, \ldots, T \qquad (1)$$

where:
$z_h^{(j)}$ – the normalized value;
$x_h^{(j)}$ – the measured value;
$X^{(j)}$ – the normalizing factor over the surveyed period (energy over analized period);
N – number of wind power farms from electrical system.

Step 4. Determination of typical load profile for wind farms: In this step, a refining of normalized characteristics occurs so that the unrepresentative characteristics are eliminated. Further, the TLP for each pattern is obtained using an averaging process of the hourly values. TLPs obtained can characterize very well the operation mode of the wind power farms, regarding to the electrical energy consumption.

5. Assignation: Finally, for each pattern of wind power farms, a TLP can be assigned.

III. Optimization Model

The objective of the planning formulation is to enhance the performance of the systems by minimize the active power losses. The main goal of the paper is to determine the optimal power system operation considering wind power sources in to an electrical network, minimizing the power losses.
A mathematical expression of the problem is:

$$\min F(X) = \min[P(U)]. \qquad (2)$$

where X is a power flow solution which stores data about the location in the system and the power capacity of the generators as well as of loads; $P(U)$ represents the power losses that depend of vector U.

To minimize the power losses into a electrical network, was used relationship (6), where R_{ij} is the resistance in to branch ij, i=1...n, j=1...n, $i \neq j$, P_i and Q_i are the real and reactive power into a node i, U_n the nominal voltage and $N_{branches}$ the number of branches in the network:

$$F_1 = \sum_{i=1}^{N_{branches}} \Delta P = \sum_{i=1}^{N_{branches}} \frac{R_{ij} * (P_i^2 + Q_i^2)}{U_n^2}. \qquad (3)$$

This item should compose with constraints to obtain the proper objective function. The main constraints in the process to determine the optimal power system operation with the proposed methodology are:

1. Voltage stability:

$$U_{i\min} \le U_i \le U_{i\max}. \qquad (4)$$

where: $U_{i\min}$, $U_{i\max}$ – minimum, maximum allowable voltage level in the system at bus i; U_i – voltage level at bus i.

2. Branch thermal limits – the power over the branch ij, S_{ij}, must be less than the maximum limit admissible that can support the line $S_{ij}^{\max}$.

$$S_{ij} \le S_{ij}^{\max} \qquad (5)$$

3. Generation limits - power generated by wind power plants is included between the maximum power allowed in bus i, $P_{DGi,\max}$.

$$P_{DGi} \le P_{DGi,\max}. \qquad (6)$$

4. Constraints for reactive power:

$$Q_{i\min} \le Q_i \le Q_{i\max}. \qquad (7)$$

where: $Q_{i\min}$, $Q_{i\max}$ - minimum, maximum allowable reactive power level in the system at bus i; Q_i - reactive power level at bus i.

5. The power losses after installing wind power plants in electrical network should be less than power losses before installing it.

$$\Delta P_{withDG} \le \Delta P_{withoutDG}. \qquad (8)$$

IV. Case Study

The proposed method was used for the optimal operation of a test electric network 220/110 kV with 10 nodes (3 nodes by 220 kV, and 7 nodes by 110 kV) and 12 branches (2, 220 kV electric lines; 7, 110 kV electric lines and 3, 220/110 kV power autotransformers), Fig. 3.

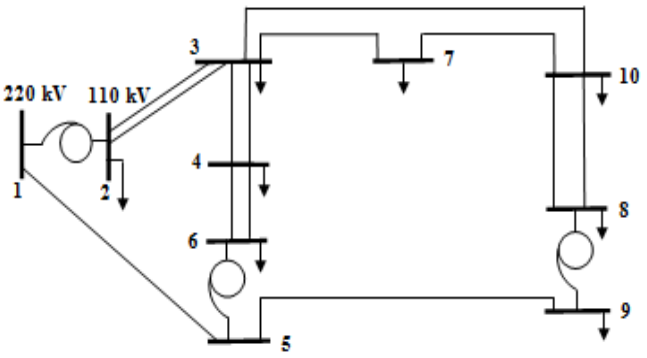


Figure 3. 220/110 kV test system.

In the test network a great potential based on wind energy was located in the region around the nodes 6 and 8. Thus, the wind power farm located in node 6 have a total installed capacity of 50 MW and the wind power farm placed in node 8, 25 MW.

In the case study a database described by generated power models, for the autumn season (3 months), corresponding to a group of wind power farms from the test electric network was considered. Every generated power model is described by 24 hourly points that depict the behavior of a wind power plant during a day.

The general information used in the clustering process concerns on generated power by the wind power plant, hour per hour for 3 months.

For determination of the optimal number of patterns, the algorithm presented in [21] was used. Thus, in the first step, the maximum number of patterns K_{max} must be determined. The

value of K_{tar} was calculated with relation $K_{max} = \sqrt{N}$, where N represents the total number of characteristics from database ($N = 90$).

In the second step, the k-means clustering method with values for K between 2 and K_{max} is used.

In the step three, the quality of grouping is evaluated using the silhouette global (SG) coefficient, Fig. 4.

Figure 4. Variation of SG coefficient

From Fig. 4, it can observe that optimum value for K is 3. For this value, in Fig. 5 is represented the forms (silhouettes) of patterns.

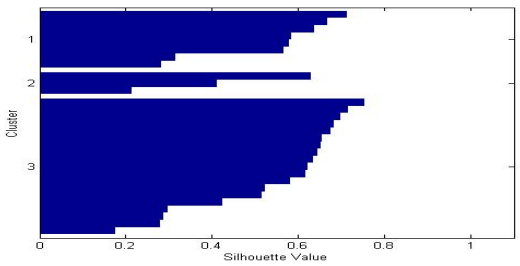


Figure 5. The forms (silhouettes) of patterns (case $K = 3$).

Each pattern is characterized by a TLP that was obtained by an averaging process of the hourly values. The wind generated power characteristics for each pattern are presented in Figs. 6-8. TLPs of the wind power farms corresponding to three obtained patterns (WPP1, WPP2, and WPP3) are indicated in Figs. 9-11.

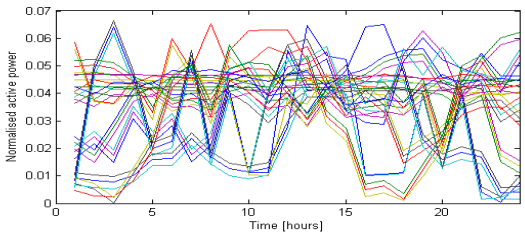


Figure 6. Generated power characteristics for WPP1pattern

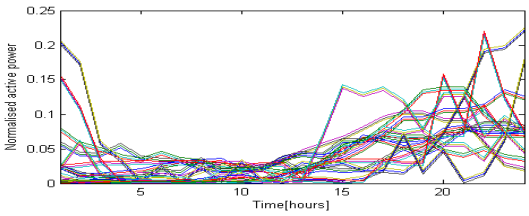


Figure 7. Generated power characteristics for WPP2 pattern

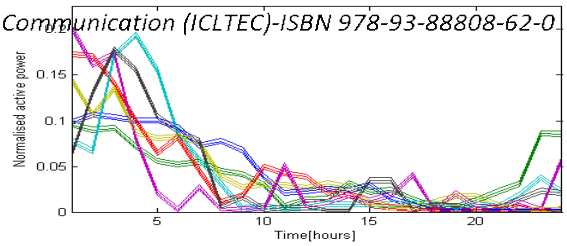


Figure 8. Generated power characteristics for WPP3 pattern

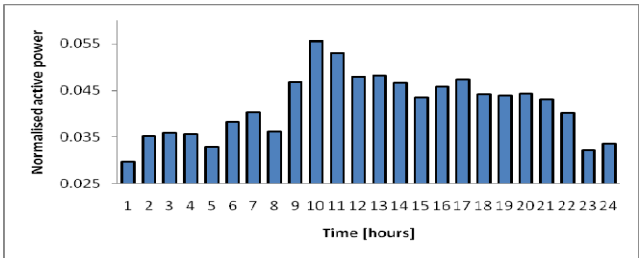


Figure 9. Typical load profile for WPP1 pattern

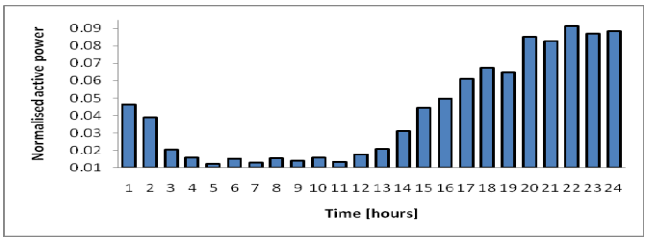


Figure 10. Typical load profile for WPP2 pattern

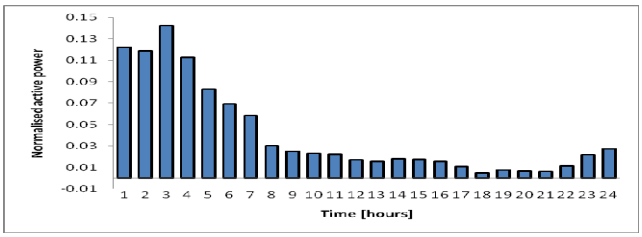


Figure 11. Typical load profile for WPP3 pattern

Further, considering these three TLPs are performed power flow calculations to analyze the evolution of the objective function. In this process the operating autotransformers plot was considered constant.

Fig. 12 shows the evolutions of the objective function in the 220/110 kV test network in the initial case (without wind power plants), in comparison with the cases when wind energy is injected coresponding to the three typical load profiles WPP1,WPP2 and WPP3.

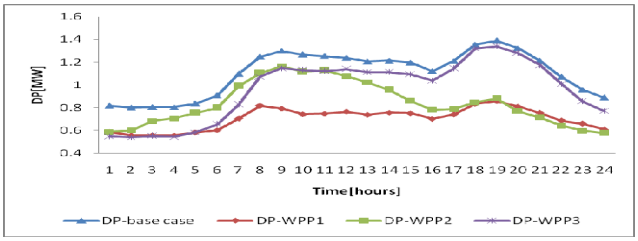


Figure 12. Objective function evolution in base case compared with cases WPP1, WPP2 and WPP3

Analyzing the results, it can see that a classification of the wind operation characteristics is useful to view the optimal operation and planning of an electric power system, on minimizing the power losses. The typical load profile WPP1 corresponds to the best power system operation for the test network, taking into consideration that has a full load operation state all the day. The objective function values, at peak load, in the base case and in case WPP1 and the voltage values evolution in nodes with wind energy injection are presented in Table I.

TABLE I. OBJECTIVE FUNCTION AT PEAK LOAD VERSUS VOLTAGE VALUES IN NODES WITH WIND ENERGY INJECTION

Objective Function				Voltage Values [kV]	
	DP[MW]	DP[%]		Base case	Case WPP1
DP-base case	1.3882	0.7744	Node 6	116.189	116.825
DP-WPP1	0.8584	0.4803	Node 8	113.674	114.297

CONCLUSIONS

In this paper a clustering technique based approach was proposed for determination the TLPs using a database described by wind power operation models, for the autumn season (3 months), corresponding to a group of wind power plants from the test electric network. The TLPs resulted describes very well the operation states of the wind power farms. So, WPP1 profile is characteristic for the operation at full load and WPP2 and WPP3 correspond to partial load operation state.

The results obtained demonstrate that the proposed approach can be used with success in the optimal power system operation on minimizing the power losses and in the same time to improve the voltage magnitude into an electric network.

REFERENCES

[1] H. Püttgen, PR. Macgregor and FC. Lambert, "Distributed generation: semantic hype or the dawn of a New Era?," IEEE Power Energy Mag., vol. 1, pp. 22-29, 2003.

[2] AG. Tsikalakis, ND. Hatziargyriou, "Environmental benefits of distributed generation with and without emissions trading," Energy Policy, vol. 35, No. 6, pp. 3395–3409, 2007.

[3] K. Alanne, A. Saari, "Distributed energy generation and sustainable development," Renewable & Sustainable Energy Reviews,vol. 10, pp. 539–558, 2006.

[4] V. Vittal, "The Impact of Renewable Resources on the Performance and Reliability of the Electricity Grid," Spring Issue of the Bridge on the Electricity Grid, vol. 40, no. 1, 2010. (https://www.nae.edu /Publications/Bridge/TheElectricityGrid/18587.aspx)

[5] A. Zervos, C. Lins and J. Muth, "A 100% Renewable Energy Vision for the European Union – Re-thinking 2050," http://www.rethinking 2050.eu/fileadmin/docum ents/ReThinking 2050_full_version_final.pdf, 2010.

[6] Renewable Energy Country Attractiveness Indices", Ernst & Young, 29 nov. 2012.

[7] Potențialul de energie regenerabilă în România, Gazeta de Agricultura, Categoria: Energie Regenerabila, http://gazetadeagricultura.info/ecobio/ 565-energie-regenerabila/11387-energie-regenerabila-in-romania.html, April 2011.

[8] Ö.S. Mutlu, "Evaluating The Impacts Of Wind Farms On Power System Operation," Journal of Naval Science and Engineering, vol. 6 , no.2, pp. 166-185, 2010.

[9] I.H. Yu, J. Lee, J.M. Ko, S.I. Kim, "A method for classification of electricity demands using load profile data," Proc.of the 4th Annual ACIS Int. Conf. Computer and Information Science, pp. 164 -168, 2005.

[10] H. Holttinen, P. Meibom, A. Orths, F. Hulle, B. Lange, M. O'Malley, J. Pierik, B. Ummels, J.O. Tande, A. Estanqueiro, M.Matos, E. Gomez, L. Söder, G. Strbac, A. Shakoor, R. João, J.C. Smith, M. Milligan and E. Ela, "Design and operation of power systems with large amounts of wind power," Final report, IEA WIND Task 25, 2009.

[11] Z. Lubosny, Wind Turbine Operation in Electric Power Systems. Berlin :Springer-Verlag, 2003.

[12] C.G. Carter-Brown, "Load profile modeling for integrated energy planning," Proc. of Domestic Use of Electrical Energy Conference, pp. 13-18, 1999.

[13] K.L. Lo, Z. Zakaria, "Electricity consumer classification using artificial intelligence," Proc. of the 39th Int. Universities Power Engineering Conf., Bristol, United Kingdom, 2004.

[14] D. Gerbec, S. Gašperič, I. Šmon, F. Gubina, "Determining the load profiles of consumers based on fuzzy logic and probability neural networks," in IEE Proc. Generation, Transmission and Distribution, vol. 151, pp. 95–400, 2004.

[15] M. Sarlak, T. Ebrahimi, S.S Karimi Madahi, "Enhancement the accuracy of daily and hourly short time load forecasting using neural network," Journal of Basic and Applied Scientific Research; vol. 2, no. 1, pp. 247-255, 2012.

[16] A.H. Nizar, "Load profiling and data mining techniques in electricity deregulated market," Power Engineering Society General Meeting, IEEE/PES, 2006.

[17] G. Chicco, "Overview and performance assessment of the clustering methods for electrical load pattern grouping," Energy, vol. 42, no. 1, pp. 68-80, 2012.

[18] N. Mahmoudi-Kohan, M.P. Moghaddam, S.M. Bidaki, "Evaluating performance of WFA K-means and Modified Follow the leader methods for clustering load curves," Proc. IEEE/PES Power System Conference and Exposition, 2009.

[19] Gh. Grigoras, M. Istrate, Fl. Scarlatache, "Electrical energy consumption estimation in water distribution systems using a clustering based method," 5th International Conference Electronics, Computers and Artificial Intelligence, Pitesti, Romania, 2013.

[20] L. R. Garcia-Escudero, A. Gordaliza, "A proposal for robust curve clustering," Journal of Classification, vol. 22, pp. 185-201, 2005.

[21] Fl. Scarlatache, Gh. Grigoraş, G. Chicco, G. Cârţină, "Using k-means clustering method in determination of the optimal placement of distributed generation sources in electrical distribution systems", International Conference on Optimization of Electrical and Electronic Equipment (OPTIM), pp. 953-958, 2012.

Design & Implementation of Efficient Multiplier Using Fixed Width RPR

M. Shiva Kumar[1], *Asst. Prof, Dept of ECE, Malla Reddy College of Engineering*
A. Suresh Reddy[2], *Asst. Prof, Dept of ECE, Malla Reddy College of Engineering*

Abstract

A reliable low area efficient multiplier is designed in this paper by using algorithmic noise tolerant architecture. The ANT architecture can achieve the demand of low power, high precision and area efficiency. ANT architecture contains a main digital signal processing block along with an error correction block. In error correction block a replica of main DSP block with reduced operands named as RPR block is used for error detection. Here different multipliers are used in main DSP block to check which performs well in ANT Architecture. The multipliers used in this paper are Baugh Wooley multiplier, Wallace Tree multiplier, Row Bypassing multiplier and Bypassing multiplier (Row and Column Bypassing multiplier).

Key words: **ANT Architecture, RPR.**

I. Introduction

In VLSI lowering the power and area of systems is the main aim. To lower the power dissipation voltage over scaling is in the process because CMOS circuit power is directly proportional to the square of supply voltage. This voltage over scaling leads to the reduction of signal to noise ratio (SNR). In algorithmic noise tolerant architecture voltage over scaling can be used to reduce the power but signal to noise ratio is maintained because of error correction block. Previously full width RPR is used in error correction block but to reduce complexity full width RPR is replaced with fixed width RPR. Using the fixed width RPR computation error can be occurred. A compensation circuit must be added to the fixed width RPR to reduce the computation error. By taking the use of probability, statistics and partial product weight analysis, an approximate compensation vector is found.

II. ANT Architecture:

There are two blocks present in the ANT architecture. One is main digital signal processing block another one is error correction block. Error correction block contains a reduced precision replica block along with the circuitry which checks the error occurred because of voltage over scaling in main digital signal processing block. The ANT architecture is as shown in fig.1.

Here the output of main block is referred as $y_a[n]$, output of RPR block is referred as $y_r[n]$ and output of error correction block is noted as $\hat{Y}[n]$.

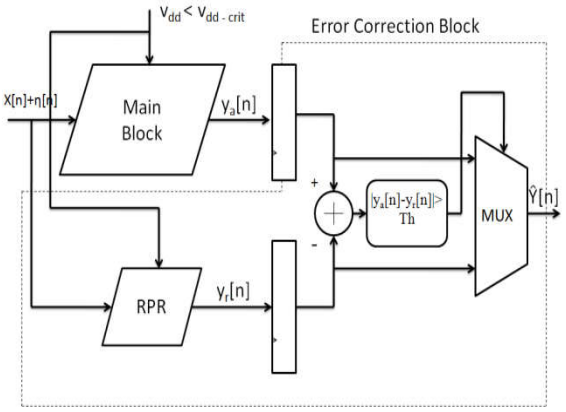


Fig1.ANT Architecture

RPR is a replica of main block with reduced precision operands and have shorter computation delay. Here fixed width RPR is taken instead of full width RPR to avoid infinite growth of bit width. If any soft errors occurred in main DSP block output $y_a[n]$, RPR output $y_r[n]$ is still correct since the critical path delay of replica is smaller than $T_{samp.}$ So $y_r[n]$ is used to detect errors in MDSP output by comparing the difference of $|y_a[n]-y_r[n]|$ against a threshold value Th. If the difference between $y_a[n]$ and $y_r[n]$ is larger than the threshold value Th, the output $\hat{Y}[n]$ is $y_r[n]$ else the output is $y_a[n]$. So $\hat{Y}[n]$ is expressed as

$$\hat{Y}[n]= y_a[n], \text{ if } |y_a[n]-y_r[n]| \leq Th$$

$$y_r[n], \text{ if } |y_a[n]-y_r[n]| > Th$$

The threshold value Th is determined as

$$Th= max_{\forall input} |y_0[n]-y_r[n]|$$

Where $y_o[n]$ is referred as error free output signal.

A full width (n/2) bit RPR can be divided into four subsets, which are most significant part (MSP), input correction vector (ICV), minor input correction vector (MICV) and least significant part (LSP). In the fixed width RPR only MSP part is kept and other parts are removed i.e. ICV, MICV, and LSP parts are truncated.

In the consideration of fixed width RPR a disadvantage is also present that there is a truncation error occurred because of the avoidance of ICV, MICV, LSB bits. To reduce the error, compensation circuit can be added to the fixed width RPR. The bits which are having highest weight in the truncated part are added as compensation circuit. Here ICV and MICV parts are used in compensation circuit because of their highest weighing.

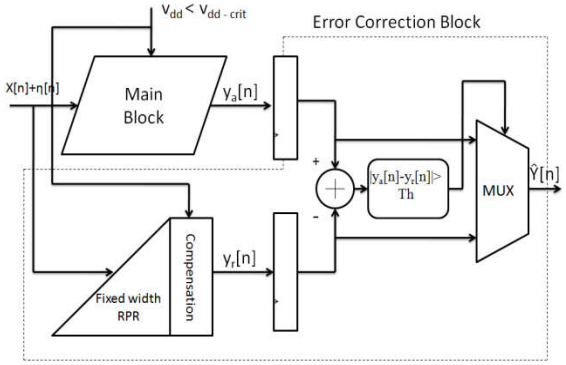


Fig 2: ANT Architecture with RPR compensation circuit

III. Baugh Wooley Multiplier

Baugh Wooley multiplier is used in the Main DSP block. Consider two unsigned inputs X and Y which are expressed as

$$X=\sum_{i=0}^{n-1} x_i\, 2^i \quad Y=\sum_{j=0}^{n-1} y_j\, 2^j$$

The result of multiplication can be expressed as

$$P=\sum_{k=0}^{2n-1} p_k\, 2^k=\sum_{j=0}^{n-1}\sum_{i=0}^{n-1} x_i\, y_j 2^{i+j} \qquad (1)$$

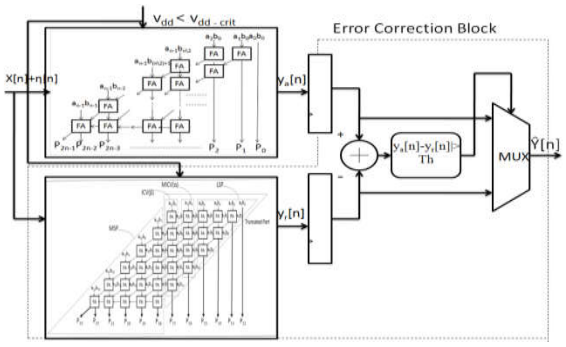


Fig 3: Baugh Wooley Multiplier in ANT architecture

For higher accuracy the error compensation circuit in the fixed width RPR can be set as shown in fig4.

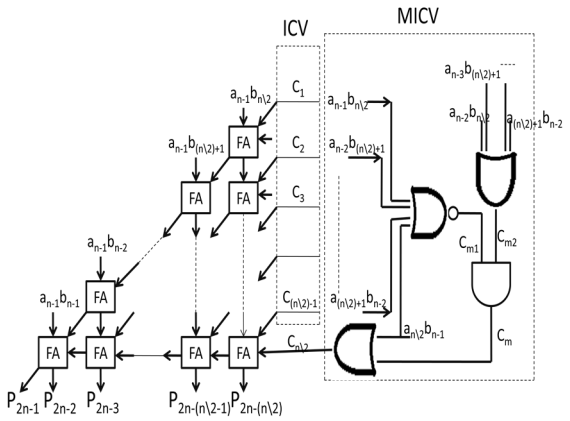


Fig 4: High accuracy fixed width RPR with compensation circuit constructed by ICV and MICV.

The output of fixed width multiplier P_t can be expressed as

$$P_t=\sum_{j=\left(\frac{n}{2}\right)+1}^{n-1}\sum_{i=\left(\frac{3n}{2}\right)-j}^{n-1} x_i\, y_j 2^{i+j} + f(\mathrm{EC})$$

Where Error correction EC is expressed as

$$f(\mathrm{EC})=f(\mathrm{ICV})+f(\mathrm{MICV})$$

IV. Row Bypass Multiplier

Generally a conventional full adder has three inputs and two outputs. When the operand bit of multiplier is zero the full adder has disadvantages such as low operational speed and unwanted switching activity. When zero partial products are added, a large number of signal transitions are generated and do not affect the final product. By using Row Bypassing multiplier zero partial products can be bypassed to achieve optimization. A modified full adder is used to achieve this optimization. The modified full adder is as in fig 5.

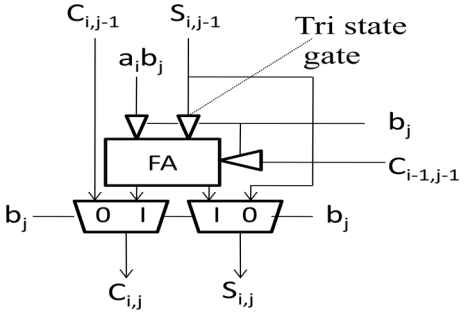


Fig 5: Structure of Full adder for Row Bypassing multiplier

In the design of n bit Row Bypass multiplier (n-1)x(n-1) full adders, 2x(n-1)x(n-1) multiplexers and 3x(n-1)x(n-1) three state gates are presented. A 4 bit Row Bypassing multiplier is as shown in fig 6.

A 12 bit Row Bypass multiplier can be designed and put in ANT architecture in the place of Baugh Wooley multiplier and performance was observed.

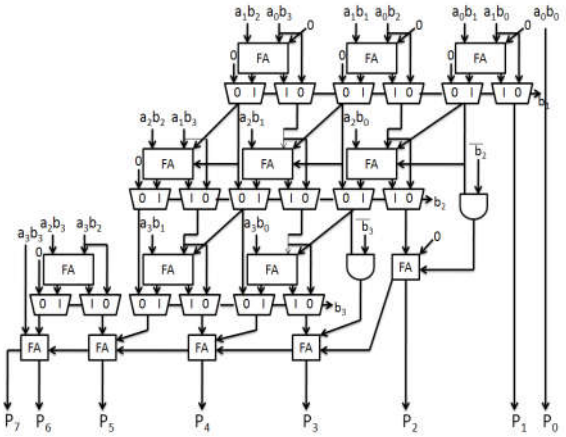


Fig 6: Row Bypass multiplier

V. Row and Column Bypassing Multiplier

Row and Column Bypassing multiplier is based on two dimensional bypassing features. According to this the addition operations in $(i+1)^{th}$ column or j^{th} row can be bypassed if the bit in that corresponding column or row is zero. The addition operation in the $(i+1,j)^{th}$ adder can be bypassed if the product a_ib_j is 0 and the carry bit $c_{i,j-1}$ is 0. If the product bit a_ib_j or the carry bit $c_{i,j-1}$ is 1, the addition operation in that corresponding full adder will be executed. It reduces the power more than that of row bypassing multiplier. The circuitry used to build this multiplier is changed as shown in fig 7(a) and 7(b).

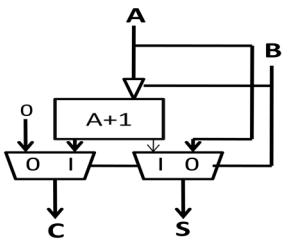


Fig 7(a): Half adder in Bypassing based design.

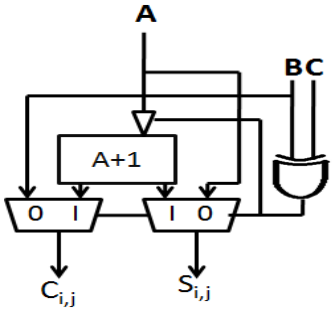


Fig 7(b): Full adder in Bypassing based design.

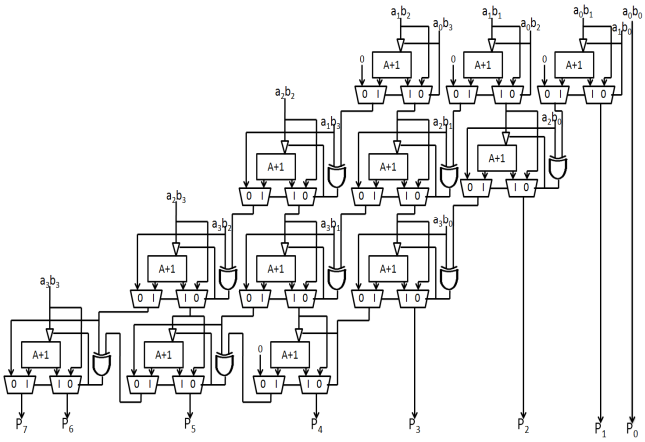


Fig 7(c): Row and Column Bypassing Multiplier

A 4 bit Row & Column bypassing multiplier is as shown in fig 7(c). A 12 X 12 bit Row & Column Bypassing multiplier is replaced in place of Baugh Wooley multiplier in the ANT architecture to check its performance.

VI. Simulation Results

These are the simulation results of ANT architecture when 12 bit Baugh Wooley multiplier, 12 bit Row Bypass multiplier and 12 bit Row and Column Bypassing multiplier are used.

Fig 8: simulation result of ANT architecture when Baugh Wooley multiplier is used.

Fig 9: simulation result of ANT architecture when Row Bypassing multiplier is used.

Fig 10: Simulation result of ANT architecture when Row and Column Bypassing multiplier is used.

VII. Conclusion

In the fixed width RPR based ANT design 12 bit Baugh Wooley multiplier, Row Bypassing multiplier and Row & Column bypassing multipliers are presented to check the area and delay performances.

The area and delay performance of Row & Column Bypassing multiplier is better than that of Baugh Wooley and Row Bypass multipliers.

The delay of Row & Column bypassing multiplieris 10% lower than Baugh Wooley multiplier and 7% less than that of Row Bypass multiplier. The area of Row & Column bypassing is 2% less than that of Baugh Wooley multiplier and 6% less than that of Row Bypass multiplier

References:

[1] I-chyn Wey, Chien-Chang Peng and Feng-Yu Liao " Reliable low power multiplier design using fixed width replica redundancy block " IEEE transactions on very large scale integration systems, vol.23, no.1, January 2015.

[2] B.Shim, S. Sridhara, and N.R.Shanbhag, "Reliable low power digital signal processing via reduced precision redundancy" IEEE Transactions on very large scale integration systems, vol.12, no.5, May 2004.

[3] Optimized Multiplier Using Bypass Technique" The International Journal of Computer Science & Applications, Volume1, no.2 April 2012.

[4] 'Low cost Low Power Improved Bypassing based Multiplier' A special issue on Recent Advances in Engineering & Technology 2013(NCRAET-2013).

[5] I.C.Wey and C.C.Wang, "Low error and area efficient fixed width multiplier by using minor input correction vector" in Proc. IEEE Int. Conf. Electron, Inf. Eng, vol.1, Kyoto, Japan, Aug 2010, pp.118-122.

[6] A.G.M.Strollo, N.Petra and D.D.Caro, "Dual tree error compensation for high performance fixed width multipliers" IEEE Trans. Circuits Syst II, Exp. Briefs, vol.52, no.8, pp 501-507 Aug 2005.

[7] "Area Efficient Low Error Compensation multiplier design using fixed width RPR" International Journal of Engineering and Technology e-ISSN:2395-0056 Volume: 02 Issue: 07 Oct-2015.

[8] "Design of low power fixed width multiplier using row bypassing" Article in IEICE Electronics Express, October 2012.

[9] J.M.Jou, S.R.Kuang and R.D.Chen "Design of low error fixed width multipliers for DSP applications" IEEE Trans, Circuits Syst, vol 46, no.6, pp. 836-842, Jun.1999.

[10] S.S.Kidambi, F.EI-Guibaly, and A.Antonou, "Area efficient multipliers for digital signal processing applications" IEEE Trans. Circuits Syst. II, Exp.Briefs, vol.43, no.2, pp.90-95, Feb. 1996.

[11] Y.C.Lim "Signal- precision multiplier with reduced circuit complexity for signal processing applications" IEEE Trans. Comput, vol.41, no.10, pp.1333-1336.

[12] "Reduction of Hardware Complexity and Truncation Error by using Fixed Width Baugh Wooley multiplier" Special Issue of International Journal of Computer Applications (0975-8887) on International Conference on Electronics, Communication and Information Systems (ICECI-12).

Modern Technology in Solar Energy Generation

Nimmala Mahesh[1],P. Manikya rao [2]
Dr. M.G.R Educational and Research Institute, [1,] Mallareddy college of Engineering[2]

Abstract:**Solar energy is playing a pivotal role in compensating the electrical energy as there is short fall in this engergy due to more demand and decline trends of conventional source of energies exhaustion of fuels like coal, petroleum, natural gases and constant of environmental and climatic changes to cope up this photovoltaic installation is being done in an electrical system to compensate and enhance the energy. an photovoltaic installation in an electrical system is made from the assembly of various photovoltaic units that uses solar energy to produce the electricity in a cheaper way from sun power. Till now the use and scope of solar energy is limited and has not reached upto masses Moreover the efficiency of the system is also low due to which the output is not sufficient as compared to input as in some installed case of solar panel it has been observed that efficiency is not more that 27%. To make it versatile and more useful for the masses newer trends and innovations will help. These have discussed in this paper.**

Keywords:**Solar panels, Efficiency, Namadd, renewable engergy resources, distributed generation**

I. Introduction

Now-a-days distributed generation (DG) is not a new concept. Without this the rectification of energy crises is not possible [1]. In most of the countries the electrical power demand is more than the electrical power generated. On the other hand there is a serious decline in the availability of natural resources, fuels, coal and gases etc. The generation of hydro power plant also varies due to variation in inflow of water from catchment area. When the capacity of hydro power plant decreases the power shortage arises. The solar power plant may be installed in such a fashion that these may work in unison for example when draught is more showering of sun is more. In this way shortage of power will be compensated by the energy governed by the solar power plant. Moreover this installation is to be done in such a way that solar panel will cover the rivers or reservoir reducing the evaporation which enhancing the capacity of dam. On the other hand the solar panel covering this area will generate electrical power which will the enhance the power generation of the system. Moreover by using some newer technologies this power generated may be integrated with the power grid to enhance the capacity of grid.

The development and uses of solar energy at large scale is not only reasonable method of energy resources utilization in the future but also effective frame to improving energy resource crises economically. There is different solar resource in different area, seasons, and weather conditions and so on because of so many influencing factors [2]. As this of kind of renewable energy is available in abundance in nature. The research and application of solar energy will be used to deal with alternative energy [3] – [4]. The advantages of renewable energy sources are enormous as they are free from gas emissions from few conventional energy resources which have impact on the global warming. If this generation of solar energy expedited rigorously can meet the most of the energy demand of the world. Use of Renewable energy will prove as panacea for solving the climatic and environmental problems as every sector of society is keen to solve these problems on the health ground problems. Currently, renewable energy sources install 15 percent to 20 percent of the world's total energy demand [5]. The solar energy is considered as the most promising and important renewable. It is envisaged that solar energy power plants would meet all human needs and would eventually replace the conventional power plants [6].

II. TYPES OF DISTRIBUTED GENERATION

There are different types of distributed generation according to the constructional and technical points of view as shown in Figure (1) [1].

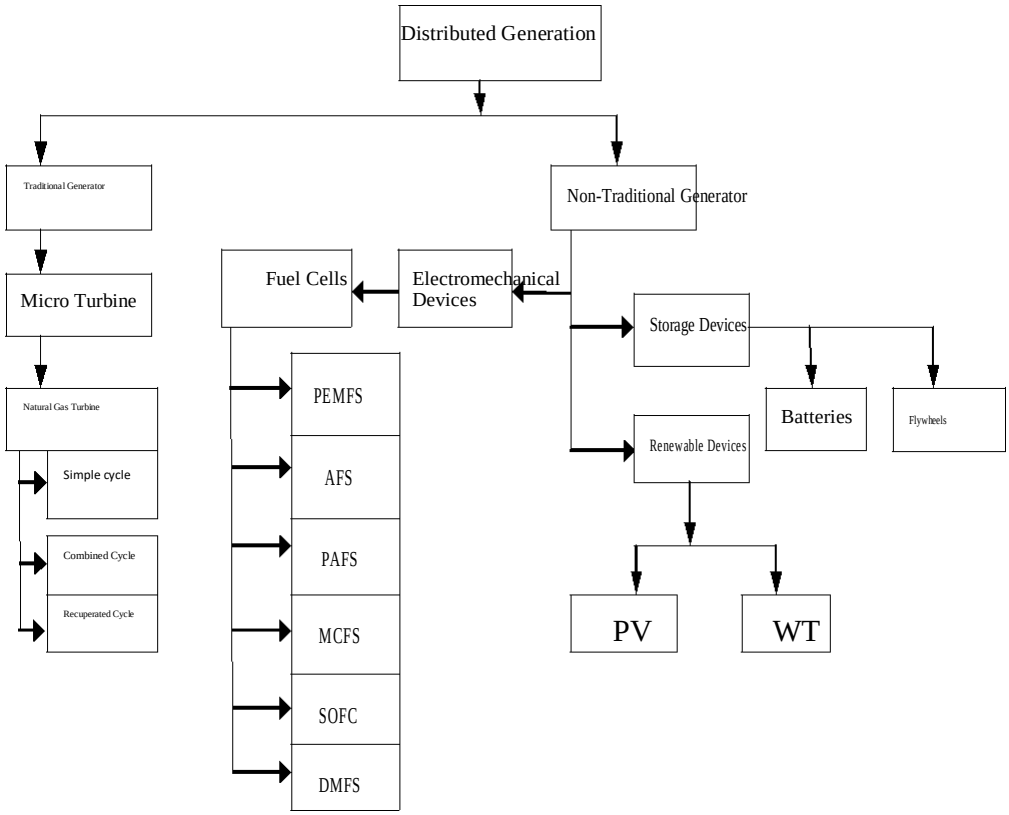


Figure 1 Types of Distributed Generation

III. The Principle Of Energy Storage

The circulation medium was heated by synchronous tracking and non-tracking solar collector and injected into the heat exchanger which was set in concrete pile with the help of circulation pump. After heat exchanging between the heat exchanger and concrete pile, circulation medium was pumped to synchronous tracking solar collector, so formed circulations.And then, the solar energy which was gathered by synchronous tracking and non-tracking solar collector was stored constantly in the underground concrete energy storage pile [7]. The circulating principle is shown in Figure(2). The solar nstal are used to generate steam which drives the steam turbines coupled with alternator as per figure (3).

IV. Solar Energy Policies Of India

The Government of India has increased its focus on developing alternativeresources of energies especially Solar Energy under the policies related to energydevelopment. The solar energy is available in abundance and almost free of cost as it is available from nature. Due to rapid economic expansion India is one of the

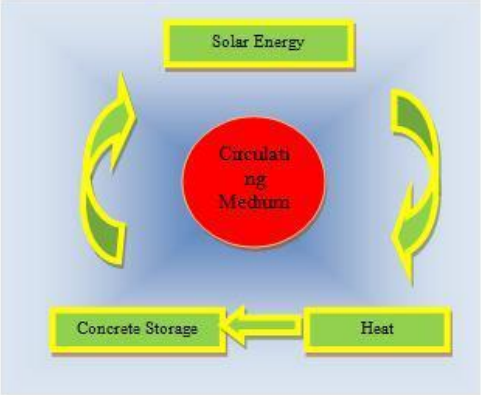


Figure 2 the circulating principle of solar energy

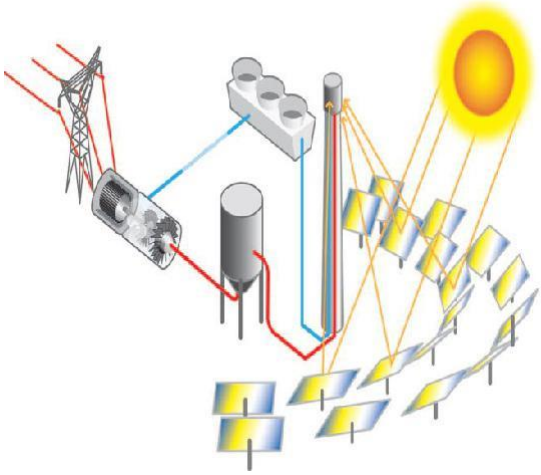


Figure 3 System of Power generation from solar energy

most growing markets and expected to be second largest energy contributor in energy market in the world by 2035. Due to limited domestic fossil fuels reserve, the India has strong planning to expand the renewable energy sources for power sector.

1. To supply the electricity to all the areas included the rural areas as mandated in section 6 of electricity act. Both the Central and State Government will jointly nstalled to achieve this objective at the earliest. Rural Electrification will be done for securing electricity access to the entire household in rural sector. Most of this requirement will be fulfilled by use of renewable energy sources.
2. Reliable rural electrification would be done either through conventional or non conventional methods of electricity whichever is more suitable and economical. Non conventional sources of energy especially Solar can be utilized even where Grid connectivity exists [8].
3. Particular attention is to be given to Dalit Bastis, Tribal areas and other weaker sections of the society the other newer resources.
4. Rural Electricity Corporation of India (REC) is the nodal agency at central govt. Level to implement these programs of electrification in rural areas. The REC will nstal all the goals set up by the National Common Minimum Programme ensuring timely implementation [9].
5. Responsibility of operation and maintenance & cost recovery could be discharged through appropriate arrangement with Panchayats, Local Authorities, BDO, and NGO etc [10].

6. This Great task of Rural Electrification requires cooperative efforts of all agencies like Govt. Of India, State Government and community education cell in rural areas.
7. The Electricity act 2003 has provision of restructuring the electricity industry which unbundled the vertically integrated electricity supply in each state. Now generation, transmission and distribution companies have been formed by the Regulatory Commission of state electricity board. Regulatory Commission will also specify the minimum percentage of electricity that each distribution utility must get from renewable energy sources [11].

V. Newer Methods Which Will Enhance the Use of Solar Energies

Day by day new trends and innovations are being developed throughout the world in R&D centers, automobile sectors and domestic use in institutions, hostels to reduce the energy wastage and to generate the power by solar devices. Many of them are explained given below:

1. Solar cells of higher efficiency have been developed having conversion efficiency more than 37% as compared to the previous solar cells having efficiency of 27 % made of two materials. Tata power is going to install the solar panels having 35% efficiency [12]. In These cells three photo absorption layers are stacked together. This has been developed by stacking Indium, Gallium and arsenide as the bottom layers. These cells have capability of absorbing the light from various wavelengths available in sunlight and convert into electrical energy. Through optimal process the active area has been increased. This breakthrough in technology has been done by new energy and industrial technology development organization.

1. The conversion efficiency of solar panels/plates is increased by newer devices of cleaning these panels. The device makes use of automated "dry-sweep'' to push dust and dirt away from the surface of these devices. In south Arabian language the device is known as nsta which is very rugged and have low maintenance cost. The device is powered by the lithium ion batteries. These batteries are charged by the array itself and have high efficiency. The device has moving parts. It is very interesting that this act like a robotic arm and automated work with scheduling. The device can jump the obstacle between the panels[13].

2. Throughout the world research and development is done to procure more and more energy from various devices and technologies. Under this concept Japan developed a fabric which is known as a solar cell fabric capable of harnessing the energy from sunlight while you are moving by wearing this fabric made cloth. This fabric is made from wafer thin solar cells woven in a stylish way. The electricity generated will be capable to charge the mobile and other portable electronic gadgets. The thread used will be stronger and which increases the life or durability of the fabric cloth. The same idea can be embedded/used in the blind makers and certain type of curtain will also generate power, when sun rayeson these. Various companies developing this type of fabric in association with solar cell maker. This will help the men to recharge these small gadgets while in sun [14].

3. A new trend of solar panel roofs have come in which most of the buildings the roof is covered with solar panel. In some advanced countries like china the roof of the max. Of the homes is made of solar panel by using aaluminum or strong alloy to support the weight of panels. In remote area where grid supply is not viable these panels generate power for themselves and supply electricity to the neighbours also who cannot afford the cost of installation. This will help nearby masses and community in that region where distribution of power is not feasible by other ways & transmission may not be possible due to heavy expenditure.

4. In some countries in urban area some hobbyist of solar energy generate electrical power this system for sufficient for their requirement and surplus generated power is supplied to the grid empowering the national grid. The solar panel should be installed on the vehicles where it is possible so that charging of batteries and other devices in the vehicle may be done with the help of solar energy. Whenever the solar rays fall on the panel this will improve the electrical efficiency of the vehicles. This type of experimentation and uses are already being done in Japan and in other advanced countries where conversion technologies from solar to electrical are being used frequently and sufficiently. In India also solar panels have been installed in metro railway service.

5. As we are aware of that electrical power demand is increasing and viewing the climatic concerns it is desired that renewable energy sources especially solar may be integrated to the utility grid. By using better flexibility in integration through power electronics. Harmonics can be reduced and the reactive power can be balanced.

6. In these days most of the power industries switchover side by side starting manufacturing of the solar inverters of high capacity. These may be utilized to get emergent power if not continues at remote locations where there is no grid supply.
7. CSP systems technology is used for power generation in the system large, flat, sunlight mirrors known as heliostats receive sun light at the top of the tower. A fluid for heat transfer is used to generation the steam which is used for production of electrical power .In some countries the capacity of these plant as high as up to 200 MW. These power tower are very popular in these days because of solar to electrical conversion efficiency is high [15].

VI. Distributed Solar Energy Generation

In [16] the authors has explained the use of Distributed solar photovoltaic (PV) systems is producing electricity onsite, so reducing the requirement to build up new transmission line and also avoiding line losses. Distributed generation also offer significant benefits to the consumers while providing resiliency to an electric grid that is based on the traditional and centralized model. These systems are used in applications ranging from small commercial to residential and for industrial use. Though this market is still primarily driven by government incentives, distributed solar PV will continue its steady march in future. Due to reduced market activity in Italy and Germany, global distributed solar photovoltaic market contracted slightly in 2012, However, growth in the United States, China, Japan, and other countries continued, driven by solar PV module price reductions, the growth of third-party financing models, and feed-in tariffs. Navigant Research forecasts that, from 2013 to 2018, 220 GW of distributed solar PV will be installed worldwide, representing $540.3 billion in revenue.

VII. Conclusion

Due to decline aviability of natural's fuels and viewing environmental changes causes due to conventional method of generation, the use ogf solar energy is becoming popular and urgency of the day. This will create healthy environment for the humain beings which are suffering from the various hazards due to pollution from the nstalled contents. Moreover the power generation due to hydro power plant is not also regular due to irrugalar flow of water from the catchment area. So it is concluded that solar power plant may be nstalled in such a way so these may work in unison with hydro and other methods of generation to enhance the clean and green energy.

References

[1] W El-Khattam and M.M.A Salama, "Distributed generation technologies, definitions and benefits", Electric Power Systems Research, vol. 71, Issue 2, p.p.,119–128, October 2004.
[2] Mingzhi Zhao, Zhizhang Liu and Mingjun Yu, "Data acquisition and analyzing of solar energy resource", Information and Automation (ICIA), Harbin, p.p., 2205-2208, 20-23 June 2010.
[3] Zhao Qingbo, Shan Baoguo, Situation and Outlook of Energy Demand in the World, Energy of China, 2002(02):34-36.
[4] Zehner and Ozzie, "Green Illusions", Lincoln and London: University of Nebraska Press. Pp., 331–42, 2012
[5] G.M. Shafiullah, A.M.T. Oo, D. Jarvis, .B.M.S. Ali, P. Wolfs, "Prospects of Solar Energy in Australia", Electrical and Computer Engineering (ICECE), 2010 International Conference, Dhaka p.p., 350 – 353, 18-20 Dec. 2010.
[6] S. Rao and B.B.Parulekar, "Energy Technology Non-Conventional, Renewable & Conventional", Khanna Publication, 3rd, 2012.
[7] Yan Zhao, Dajun Zhao, Yirong Yao, Xin Fang ,Shousheng Li, "Development of Solar Energy Underground Seasonal Storage Device and Its Parameters Measuring System", Information Engineering (ICIE), 2010 WASE International Conference, Beidaihe, Hebei, p.p., 231 – 234, 14-15 Aug. 2010.
[8] G. D. Kamalapur and R. Y. Udaykumar, "Rural Electrification in the Changing Paradigm of Power Sector Reforms in India", International Journal of Electrical and Computer Engineering, p.p., 147-154, vol. 2, No.2, April 2012.
[9] N. Sasidhar, "Electricity online trading in India", Rural Electricity Corporation of India, 2012.
[10] Giuseppe Buglione, Guido Cervigni, Eileen Fumagalli, Elena Fumagalli and Clara Poletti, "Integrating European Electricity Markets", Center for Research on energy and environmental economics policy, Report 2, Oct. 2009.
[11] MadhuKhanna, KusumMundra, AmanUllah, "Parametric and Semi-Parametric Estimation of the Effect of Firm Attributes on Efficiency: Electricity Generating Industry in India", Journal of International Trade and Economic Development. Pp. 419–436.vol. 8, No. 4, Sept. 2011.
[12] Electronics for You, "Most efficient Solar Cell", Technology news, January, 2013.

[13] GeorgeEitelhube, "Seed Fund Winners - NOMADD: The No-water Mechanical Automated Dusting Device", King Abdullah University of Science and Technology (KAUST), energy and sustainability are primary research drivers, 2012.
[14] John Badding, "Flexible Silicon Solar-Cell Fabrics May Soon Become Possible", journal Advanced Materials,Penn State University, 6 December 2012.
[15] DolfGielen, "Concentrating Solar Power technology", Energy Technology Systems Analysis Programme, International Renewable Energy Agency, January 2013.
[16] G. D. Rai, "Non-Conventional Energy Sources", Khanna Publication, 4th edition, 2012.

International Conference on Latest Trends in Electronics and Communication (ICLTEC)-ISBN 978-93-88808-62-0

Talking without Talking: A Solution to Noisy Communication

K MANIKANTA
B.Tech M.Tech(ES & VLSI)
Department of ECE
Malla Reddy College Of Engineering
Hyderabad,INDIA

M.Rani
B.Tech,M.Tech(DECS)
Department of ECE
Malla Reddy College Of Engineering
Hyderabad,INDIA

***Abstract*— when we are in movie, theatre, bus, train there is lot of noise around us we can't speak properly on a mobile phone. In future this problem is eliminated with the help of Silent sound technology. It is a technology that helps you to transmit information without using your vocal cords. This technology notices every lip movements & transforms them into a computer generated sound that can be transmitted over a phone. Hence person on another end of phone receives the information in audio. It uses electromyography, monitoring tiny muscular movements that occur when we speak and converting them into electrical pulses that can then be turned into speech, without a sound uttered .When demonstrated, it seems to detect every lip movement and internally converts the electrical pulses into sounds signals and sends them neglecting all other surrounding noise. So, basically, it reads your lips. It is definitely going to be a good solution for those feeling annoyed when other speak loud over phone.**

***Keywords*-- silent sound, electromyography, electromyograms, digital image pr**

I. INTRODUCTION

Silent sound technology enables speech communication to take place when an audible acoustic sound is unavailable.

By acquiring sensor data from elements of the human speech production process- from the articulators ,their neural pathways, or the brain itself –it produces a digital representation of

speech which can be synthesized directly, interpreted as data, or routed into a communication networks[2].

Sound Technology is a technology for mobile phones that helps you communicate in noisy places too. It is a technology that will help reduce noise pollution to a great extent. The uses of this technology are immense for people who are vocally challenged or have been rendered mute due to accident.

Humans are capable of producing and understanding whispered speech in quiet environments at Silent remarkably low signal levels. Most people can also understand a few words which are unspoken, by lip-reading The idea of interpreting silent speech electronically or with a computer has been around for a long time, and was popularized in the 1968 Stanley Kubrick science- fiction film ''2001 – A Space Odyssey " [7].A major focal point was the DARPA Advanced Speech Encoding Program (ASE) of the early 2000's, which funded research on low bit rate speech synthesis ''with acceptable intelligibility, quality , and aural speaker recognizability in acoustically harsh environments".

International Conference on Latest Trends in Electronics and Communication (ICLTEC)-ISBN 978-93-88808-62-0

Fig. 1 Many people talking at one place

II.PROCESS OF SPEAKING

The air passes through the larynx and the tongue and the words are formed with the help of the articulator muscles in the mouth and the jaw. The articulator muscles are activated irrespective of the fact that jo air passes through them or not. The weak signals are sent from the brain to the speech muscle. These signals are collectively known as the electromyograms.

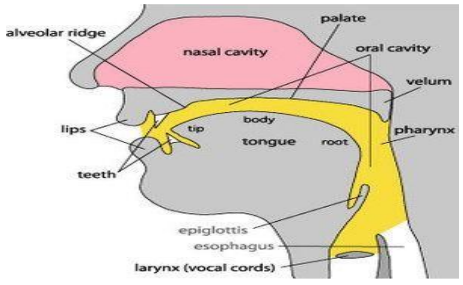


Fig. 2 Speaking Process in human body

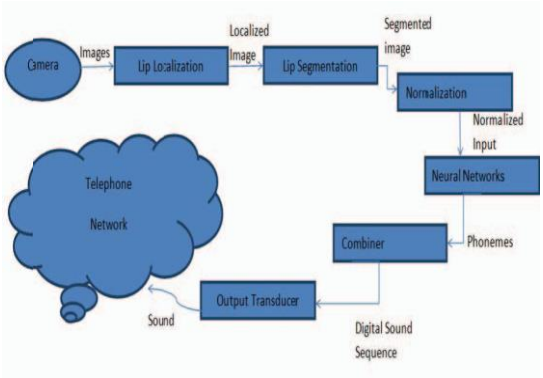


Fig. 3 Block Digram Showing different steps in generation of sound from lip movements

III.METHODS

Silent sound technology is processed in two ways

.They are

A. Electromyography (EMG)
B. Image Processing

A.Electromyography

Electromyography is a technique used in silent sound technology that monitors tiny muscular movements that occur when we speak and converting them into electrical pulses that can then be turned into speech, without a sound utter. Electromyography (EMG) is a technique for evaluating and recording the electrical activity produced by skeletal muscles.EMG is performed using instrument called an **electromyograph**, to produce a record called an **electromyogram**. An electromyograph detects the electrical potential generated by muscle cells when these cells are electrically or neurologically activated [2].

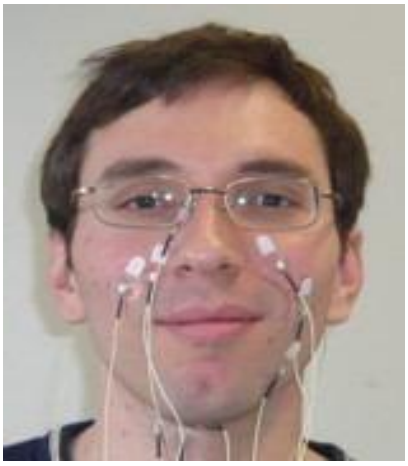


Fig. 4 Electromyographic sensors attached to face.

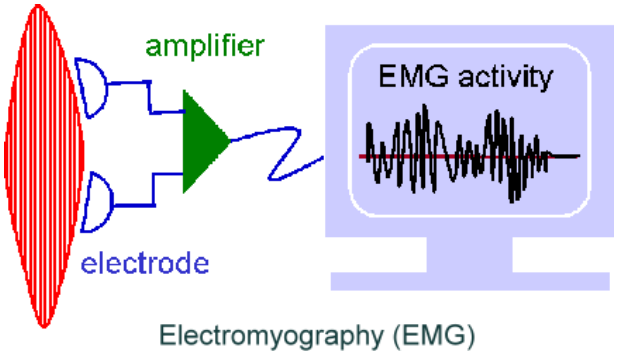


Fig. 5 Electromyography activity

International Conference on Latest Trends in Electronics and Communication (ICLTEC)-ISBN 978-93-88808-62-0

B.Image Processing

The simplest form of digital image processing converts the digital data tape into a film image with minimal corrections and calibrations. Then large mainframe computers are employed for sophisticated interactive manipulation of the data. In the present context, overhead prospective are employed to analyze the picture. In electrical engineering and computer science, image processing is any form of signal processing for which the input is an image, such as a photograph or video frame; the output of image processing may be either an image or, a set of characteristics or parameters related to the image. Most image- processing techniques involve treating the image as a two-dimensional signal and applying standard signal-processing techniques to it [1].

Analysis of remotely sensed data is done using various image processing techniques and methods that includes

1) Analog Image processing: - Analog processing technique is applied to hard copy data such as photographs or printouts. It adopts certain elements of interpretation, such as primary element, spatial arrangement etc. With the combination of multi-concept of examining remotely sensed data it allows us to make a verdict not only as to what an object is but also its importance. Apart from these it also includes optical photogrammetric techniques allowing for precise measurement of the height, width, location, etc. of an object

2)

2)Digital Image Processing: - Digital Image Processing involves a collection of techniques for the manipulation of digital images by computers. It contain some flaws. To overcome the flaws and deficiencies in order to get the originality of the data, it needs to undergo several steps of processing. Digital Image Processing undergoes three general steps: 1) Pre-processing Display 2) Enhancement 3) Information extraction

IV. ADVANTAGES OF SILENT SOUND TECHNOLOGY

- Very useful for those people who lost their voice and has been rendered mute due to accident.
- At public crowded places like in market, bus, train, malls, theater etc.
- Very good technology for noise cancellation technique.
- Helps in making phone calls in noisy environment.
- Very useful for sharing confidential information like
- secret PIN number on phone at public place.

V.RESTRICTIONS

This technology works in many languages of user's choice like English, French & German, etc. But, for the languages like Chinese is difficult because different tones can hold many different meanings.

VI.FUTURE SCOPE

Silent sound technology gives way to a bright future to speech recognition technology from simple voice commands to memorandum dictated over the phone all this is fairly possible in noisy public places. Without having electrodes hanging all around your face, these electrodes will be incorporated into cell phones. Nano technology will be a mentionable step towards making the device handy.

VII.CONCLUSIONS

Engineers claim that the device is working with 99 percent efficiency. Silent Sound Technology, one of the recent trends in the field of information technology implements 'Talking Without Talking'. It will be one of the innovation and useful technology and in mere future this technology will be use in our day to day life.

International Conference on Latest Trends in Electronics and Communication (ICLTEC)-ISBN 978-93-88808-62-0

REFERENCES

[1] Hueber T, Benaroya E-L, Chollet G, Denby B, Dreyfus G, Stone M. (2010). Development of a silent speech interface driven by ultrasound and optical images of the tongue and lips. Speech Communication.

[2] Denby B, Schultz T, Honda K, Hueber T, Gilbert J.M., Brumberg J.S. (2010). Silent speech interfaces. Speech Communication

[3] http://www.techpark.net/2010/03/04/ silent-soundtechnology-an-end-to-noisy-communications/

[4] Shehjar Safaya , Kameshwar Sharma , Silent Sound Technology-An End to Noisy Communication, Speech Communication , Vol. 1, Issue 9, November 2013

Delay Estimation Model for High speed Interconnects in VLSI

International Conference on Latest Trends in Electronics and Communication (ICTEC 2013) ISBN 978-93-82208-62-0

B.Balaji[1], Professor, Dept of ECE, Mallareddy College of Engineering
R. Purushottam Naik[2,] Professor,Dept of ECE, Mallareddy College of Engineering

Abstract: In recent days there is huge demand for high speed VLSI networks. In order to judge the behavior of on-chip interconnects the coupling capacitances and interconnect delays plays a major role. As we switch to lower technology there is on-chip inductance effect that leads to interconnect delay. In this paper we try to apply second order transfer function designed with finite difference equation and Laplace transform at the source and load termination ends. Analysis shows that current mode signaling in VLSI interconnects provides times better delay performance than voltage mode.

Keywords: Current Mode, Voltage Mode, VLSI Interconnect

I. INTRODUCTION

As the number of transistors on a chip continues to increase, on-chip communication becomes a more important facet of architectural design. Traditional electrical wires, typically driven by digital components using simplistic digital signals have issues to address in the scaling chip multiprocessor market, specifically latency and energy. Global wire latency remains relatively constant, translating to a larger relative latency for even moderately-sized systems. In order to ensure signal quality, digital repeaters and packet-switching routers must be added to facilitate the transmission of long distance communications, contributing further to the latency and energy issues. Current research focuses on a few categories of solutions, each with unique benefits and limitations. The current convention is the use of packet-switching networks topologies to provide the interconnect backbone for chip-multiprocessors. A packet-switched network-on-chip (NoC) provides in-field scalability, the ability to use commercial-off-the-shelf components, and high aggregate throughput. However, a NoC also requires higher power routers and potentially long latencies for long distance communication. Another state-of-the-art solution for interconnects uses on-chip optics. Research is currently being proposed to use either waveguides or free-space optics to provide a high-throughput, low-energy, low-latency medium for on-chip communication. On the other hand, optics also has issues that prohibit its immediate use as an interconnect backbone.

First, current optical components are not easy to integrate into standard silicon CMOS process, making it more difficult to fabricate with the current technologies without sacrificing electro-optical conversion efficiency. Additionally, while some on-chip lasers exist, most often, off-chip lasers are used to provide the optical power, shifting the onus of energy efficient operation off-chip, but not removing it from the system.

II. RELATED WORKS

Signaling in global strains is a main bottleneck in excessive performance VLSI systems because of the dominant problem of signal propagation delays in comparison to circuit delays. So, correct and accurate delay estimation models are required in cutting-edge-mode signaling. Accurate

estimation of propagation delay in worldwide interconnects plays a main place in the early design levels of VLSI systems compared to neighborhood interconnect delays because, global wires supports predominant capabilities like clock, sign distribution among the useful blocks and affords power/ground to all functions on a chip. Starting from lumped RC model to distributed RLC version, diverse techniques [6]-[8] based on analytical closed form formulations had been proposed to model delay in voltage-mode interconnects. Similarly for modern-day-mode RC interconnects, closed-form delay analysis version becomes presented in [9] and the evaluation does not encompass the fast aspect input reaction. In [10] closed-form delay model for allotted contemporary-mode RC line is provided, which included the practical rapid edge input reaction issue but this version should now not include the inductance impact in present day-mode interconnect. A delay estimation model [11], that's derived the use of the concept of soaking up inductance impact into equal RC version, then changed nodal evaluation (MNA) become used. Various closed-form delay models [9]-[13] for on-chip interconnects in present day-mode signaling have more inaccuracy in phrases of postpone estimation.

III. FINITE DIFFERENCE EQUATION AND LAPLACE TRANSFORM

The Taylor series expansion for a function of one variable about the point x is

$$f(x+h) = f(x) + h\,f'(x) + \frac{h^2}{2!}f''(x) + O(h^3)$$

…….. (1)

The notation $O(h^3)$ indicates that the series, when truncated at the quadratic term in h, contains errors that scale as h^3 and higher powers of h.

We can immediately obtain an approximation to the derivative of f(x) from the first two terms of the expansion.

$$f'(x) = \frac{f(x+h) - f(x)}{h} + O(h^1) \text{ …….. (2)}$$

Note that, even though we neglect terms of $O(h^2)$ in the expansion, since we divide through by h to obtain the derivative expression the approximation is correct to $O(h^1)$ only. Note that there is an asymmetry in this approximation to the derivative at x, since the function at x and x+h occur, but not the function at x-h. This is therefore referred to as a forward difference approximation. It is possible to expand f(x) in the negative direction in the Taylor expansion and hence to obtain a backward difference approximation

$$f(x-h) = f(x) - h\,f'(x) + \frac{h^2}{2!}f''(x) + O(h^3)$$

…….. (3)

$$f'(x) = \frac{f(x) - f(x-h)}{h} + O(h^1) \text{ …….. (4)}$$

By combining forward and backward difference approximations it is possible to obtain a central difference approximation to the derivative of f(x) at x that contains errors of order $O(h^2)$. Note the difference in scaling of errors for the central difference approximation when compared with the forward and backward difference approximations.

$$f(x+h) = f(x) + h\,f'(x) + \frac{h^2}{2!}f''(x) + O(h^3)$$

$$f(x-h) = f(x) - h\,f'(x) + \frac{h^2}{2!}f''(x) + O(h^3)$$

$$f(x+h) - f(x-h) = 2h f'(x) + 2\frac{h^3}{3!} f'''(x) + O(h^4)$$

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} + O(h^2)$$

…….. (5)

By retaining terms in the Taylor series to order h^3 we can obtain an approximation for the second derivative which contains errors of order h^2

$$f(x+h) = f(x) + h f'(x) + \frac{h^2}{2!} f''(x) + \frac{h^3}{3!} f'''(x) + O(h^4)$$

$$f(x-h) = f(x) - h f'(x) + \frac{h^2}{2!} f''(x) - \frac{h^3}{3!} f'''(x) + O(h^4)$$

…….. (6)

Add these expansions to obtain

$$h^2 f''(x) = f(x+h) - 2f(x) + f(x-h) + O(h^4)$$

$$f''(x) = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} + O(h^2)$$

…….. (7)

A. Taylor series expansions in more than one dimension

A PDE contains at least two independent variables and so we need to approximate differential operators in at least two dimensions. This is done using Taylor series expansions in more than one dimension. Suppose

u = u(x,y) …….. (8)

The Taylor series expansion of u about the point (x,y) is

$$u(x+h, y+k) = u(x, y) + h\, u_x(x, y) + k\, u_y(x, y) + \frac{h^2}{2!} u_{xx}(x, y) + \frac{2hk}{2!} u_{xy}(x, y) + \frac{k^2}{2!} u_{yy}(x, y) + O(|\mathbf{h}|^3)$$

$$\mathbf{h} = \begin{pmatrix} h \\ k \end{pmatrix}$$

…….. (9)

The vector notation for this expansion is

$$u(\mathbf{x}+\mathbf{h}) = u(\mathbf{x}) + \mathbf{h}^{\mathbf{T}} . \nabla u(\mathbf{x}) + \frac{1}{2!} \mathbf{h}^{\mathbf{T}} . \nabla\nabla u(\mathbf{x}) . \mathbf{h} + O(|\mathbf{h}|^3)$$

$$\mathbf{h} = \begin{pmatrix} h \\ k \end{pmatrix} \quad \mathbf{h}^{\mathbf{T}} =$$

$$\begin{pmatrix} h & k \end{pmatrix} \mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix} \quad \nabla\nabla = \frac{\partial^2}{\partial x_i \partial x_j}$$

…….. (10)

Returning to the long-hand notation, the expansion of u(x,y) in the x direction is

$$u(x+h, y) = u(x, y) + h\, u_x(x, y) + \frac{h^2}{2!} u_{xx}(x, y) + O(h^3)$$

$$u(x-h, y) = u(x, y) - h\, u_x(x, y) + \frac{h^2}{2!} u_{xx}(x, y) + O(h^3)$$

…….. (11)

If we subtract these two equations and rearrange to make u_x the subject of the equation we find that the central difference approximation to u_x is

$$u_x(x, y) = \frac{u(x+h, y) - u(x-h, y)}{2h} + O(h^2)$$

…….. (12)

We can also obtain forward and backward difference approximations which contain errors of order $O(h^1)$ from the Taylor series expansion in either direction. It is convenient to write the function u at points on a grid for numerical solution with subscripted indices rather than arguments. Thus we make the equivalence

$$u(x, y+k) = u_{i+1, j} \quad \text{…….. (13)}$$

k is the step size or distance between gridpoints in the y direction in the numerical solution. Perversely, Farlow reverses the order of the arguments/indices in going over to the gridpoint index. We will do the same to maintain consistency with Farlow. The indicial representation of the first and second order partial derivatives is given below and illustrated

by corresponding 'computational molecules'.

$$u_x(x, y) = \frac{1}{2h}\left(u_{i,j+1} - u_{i,j-1}\right) + O(h^2)$$

$$u_y(x, y) = \frac{1}{2k}\left(u_{i+1,j} - u_{i-1,j}\right) + O(k^2)$$

$$u_{xx}(x, y) = \frac{1}{h^2}\left(u_{i,j+1} - 2u_{i,j} + u_{i,j-1}\right) + O(h^2)$$

$$u_{yy}(x, y) = \frac{1}{k^2}\left(u_{i+1,j} - 2u_{i,j} + u_{i-1,j}\right) + O(k^2)$$

…….. (14)

Provided that the step sizes in the two directions (h and k) are equal, then we obtain the following approximation for the Laplacian operator in 2 dimensions.

$$\nabla^2 u(x, y) = \frac{1}{h^2}\left(u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 4u_{i,j}\right) + O(h^2)$$

…….. (15)

Voltage mode interconnects

Voltage mode signaling is most widely used in VLSI chips. In voltage mode signaling, receiver provides high input impedance (ideally infinity). The information is conveyed in the form of voltage. The output voltage is a function of input signal and is varied according to supply voltage. Fig.1 shows the theoretical model of conventional voltage mode interconnect implementation [5]. The output is terminated by an open circuit.

CMOS representation of voltage mode is shown in Fig. 2 [1, 3]. The driver consists of an inverter which drives long RC interconnect chain. This is terminated by high input impedance of the inverter circuit at the receiver. This high input impedance of the receiver gives rise to high input capacitance which leads to high charging and discharging time for RC interconnect chain. Hence voltage mode signaling has large delay. Due to high input impedance at the receiver, the charge accumulated at the input of the receiver does not get effective discharge path to ground as a result this may cause electrostatic induced gate oxide break down.

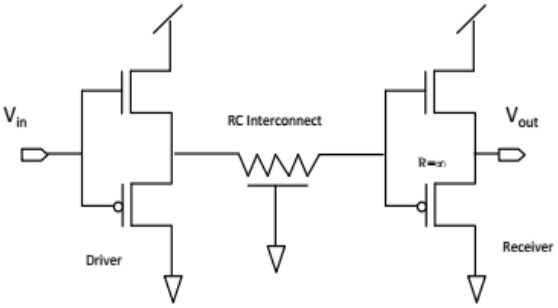


Fig. 2. CMOS representation of voltage mode signaling [5].

Current mode interconnects

In current mode signaling, information is represented as current signal. The receiver provides low impedance (ideally zero) at its input. In current mode signaling line is terminated by shorting the wire. The theoretical model of current mode signaling is as shown in Fig. 3 [5]

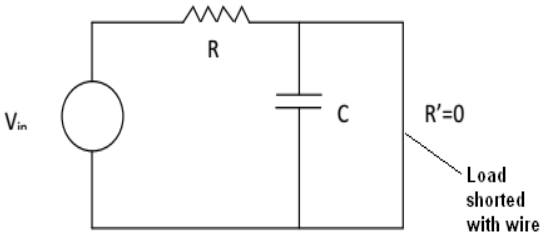


Fig. 3. Current mode signaling

The CMOS representation of current mode signaling is shown in Fig. 4. The receiver senses current signal at its input and provides low impedance.

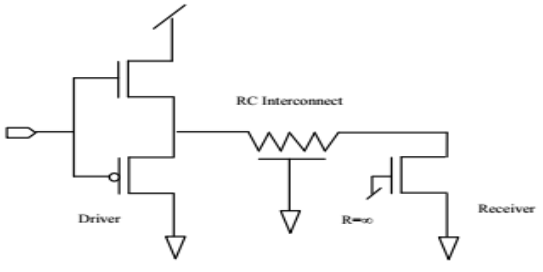
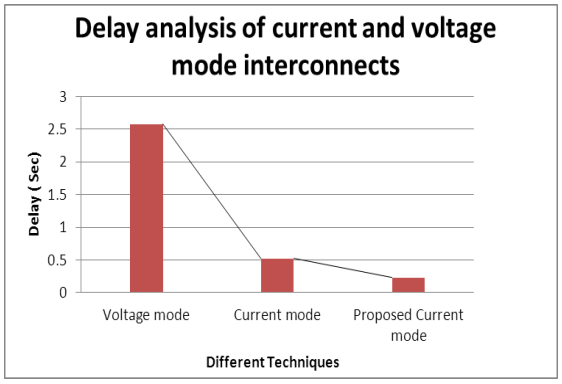


Fig. 4. CMOS representation of current mode signaling [5]

Table I Delay Analysis for voltage and Current Mode Interconnects [13, 14].

Signaling mode	Delay (ns)	P Total) (µw)
Voltage mode	2.58	22.01
Current mode	0.520	115.08
Proposed Current mode	0.010	19.33

Table I shows the delay analysis for both voltage and current mode interconnects [13-14]. It is analyzed that with Proposed Current mode interconnects delay decreases from 2.58ns to 0.010ns however power dissipation in the circuit increases from 22.01µw to 19.33µw. This is due to the low impedance at the receiver of current mode interconnect circuit. The overall figure of merit power-delay-product (PDP) of current mode increases. This displays the improved performance and advantage of current mode over voltage mode interconnects. The delay analysis for current and voltage mode interconnect is shown in Fig. 10. It is seen that there is 79.84% reduction in current mode interconnect delay.



IV. CONCLUSION

This paper presents second order transfer function designed with finite difference equation and Laplace transform at the source and load termination ends current mode signaling for delay estimation of current mode high speed VLSI interconnects. The perseverance of the current study is to estimate the delay of current-mode VLSI interconnects and to find the interaction between delay for various lengths, line inductances and load capacitances using existing voltage mode. All the benefits give current mode signaling an upper edge over the voltage mode signaling. At highly miniaturized technologies, interconnects with current mode signaling would be the best choice with the assistance of HSPICE tool.

REFERENCES

[1] V. Adler, and E.G Friedman, "Repeater design to reduce delay and power in resistive interconnect," IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, vol. 45, pp. 607-616, 1998.

[2] V. V.Deodhar, and J. A. Davis, "Voltage scaling, wire sizing and repeater insertion design rules for wave-pipelined VLSI global interconnect circuits," Sixth International Symposium on Quality of Electronic Design, pp. 592-597, 2005.

[3] R. Chandel, S. Sarkar, and R.P. Agarwal, "An Analysis of Interconnect Delay Minimization by Low-Voltage Repeater Insertion," Microelectronics Journal, Elsevier Science, vol. 38, no. 4-5, pp. 649-655, April-May 2007.

[4] R. Dhiman, and R. Chandel, "Sub Threshold Delay and Power Analysis of Complementary Metal-Oxide Semiconductor Buffer Inserted Interconnect for ultra-Low Power Applications," Journal of Low Power Electronics, ASP, vol. 8, no.1, 2012.

[5] A. Maheshwari, and W. Burleson, "Current sensing Technique for Global Interconnects in Very Deep Submicron (VDSM) CMOS," IEEE Computer Society Workshop on VLSI, pp. 66-70, 2001.

[6] W. C. Elmore, "The transient response of damped linear networks with particular regard to wideband amplifiers,"J. Appl. Phys., vol. 19, no.1, pp. 55–63, Jan. 1948.

[7] T.Sakurai, "Closed-form expressions for interconnection delay, coupling, and crosstalk in VLSI's,"IEEE Trans. Electron Devices, vol. 40, no. 1, pp. 118–124, Jan. 1993.

[8] Kahng A B, Muddu S, "An analytical delay for RLC interconnects", IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems, 1997, 16(12): 1507−1514.

[9] E. Seevinck, P. van Beers, H. Ontrop, "Current-Mohe Techniques for High-speed VLSI Circuits with Application to Current Sense Amplifier for CMOS SRAM's," IEEE J. of Solid-state Circuits, Vol 26, No. 4, April1991.

[10] Bashirullah, R.; Wentai Liu; Cavin, R., III, "Delay and power model for Current-mode signaling in deep submicron global interconnects," Custom Integrated Circuits Conference, 2002. Proceedings of the IEEE 2002, vol., no., pp.513,516, 2002.

[11] Mingcui Zhou; Wentai Liu; Sivaprakasam, M., "A closed-form delay formula for on-chip RLC interconnects in Current-mode signaling," Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on, vol., no., pp.1082, 1085 Vol. 2, 23-26 May 2005.

[12] Jaday, S.; Khanna, G.; Kumar, A.; Saini, G., "Low power high throughput Current-mode signalling technique for global VLSI interconnect," Computer and Communication Technology (ICCCT), 2010 International Conference on , vol., no., pp.290,295, 17-19 Sept. 2010.

[13] R. Bashirullah, W. Liu, and R. Cavin, "Current-mode signaling in deep submicrometer global interconnects,"IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol. 11, no. 3, pp. 406–417, Jun. 2003.

[14] International Technology Roadmap for Semiconductors (ITRS), 2009 update. [Online document], Available http: //public.itrs.net.

IOT Based Light Intensity Monitoring System using Embedded Linux & Raspberry Pi

KESANAPALLI ASKA RANI1 ,BODDU MANJULA

1.Assistant Professor, Dept of ECE, MRCE, HYDERABAD,Email:aksaranik@gmail.com.

2.Assistant Professor, Dept of ECE, MRCE,HYDERABAD , E-mail: manjulakoteswararao007@gmail.com.

Abstract: Accurate and quantifiable measurement of light is essential in creating desired outcomes in practical day to day applications as well as unique applications such as Traffic lighting system, Poultry Industry, Gardening, Museum lighting system, at emergency exits etc. Hence, Light measurement and analysis is an important step in ensuring efficiency and safety. Many of the industries are burdened with limited number of resources and real shortage of experts on their fields; real time remote monitoring presents an effective solution that minimizes their efforts and expenditures to achieve the desired results within time. This paper introduces real time remote Light intensity monitoring system using Raspberry Pi which enables the user to track the lighting system remotely. Raspberry pi is a low cost ARM11 processor Linux based computer which acts as a server, and it communicates with clients with LAN or external Wi-Fi module. The key feature of this system is light intensity being monitored instantaneously and data stored in the database for future use, and shown in the form of dynamic charts to the user according to the user requirement in a terminal device like Tablet or Smart Phone or any internet enabled device. This empowers experts to make right decisions at right time to get desired results.

Keywords: LDR. Light Intensity, Temperature, Smoke, Raspberry Pi, Web Server, Camera, Buzzer.

I.INTRODUCTION

There are many applications available to measuring and maintain the sufficient light levels such as laboratories, hospitals, educational institute, etc. To sustain healthier and safety environment enough light levels in the premises are needed. Without any distraction of whether condition, the light intensity has to be adequate for light levels intensity Some of important locations and light intensity is shown in Table I.

TABLE I. Optimum Average Light Intensity at Various Locations. Consider following Applications as an Example

Location	Illuminance(LUX)
Homes, Theaters	150
Library(Reading Area)	200
General Office work	500
Class room	300

A.Traffic Lighting System:- To ensure safety on the road, traffic lights need to be clearly visible for road users. The light intensity has to be sufficient under every (weather) condition, which set in legal standards. Over the course of time, the luminous intensity of traffic lights slowly decreases. Possible reasons are pollution of lenses or reflectors, aging of the light source

or individual LED failure. Remote monitoring enables the road authority to carry out timely services, in such a way that traffic lights keep satisfying the statutory rules for optimal traffic safety.

B. Poultry Industry:-Light Intensity is an important management factor in poultry industry to obtain optimal production. The intensity depends upon the age and type of housing being used, and type of chicken, be it broiler, breeder or layer. With blackout housing both male and female can be exposed to 3.5 fc from day one to day six and then placed on 1 fc to 19 or 20 weeks. After 19 - 20 weeks the broiler breeders can be exposed to about 3.0 to 5.0 fc during the entire production period. Layers should be exposed to about .5 to 1.5 fc (One foot-candle = 10.76 lux) for better production [4-6].

C. Plants Growth:-Deficient light intensity tend to reduce plant growth, development and yield. This is because low amount of solar energy restricts the rate of photosynthesis. Below a minimum intensity, the plant falls below the compensation point. Compensation point is the metabolic point at which the rates of photosynthesis and respiration are equal so that leaves do not gain or lose dry matter. Photosynthesis significantly slows down or ceases while respiration continues. Likewise, excessive light intensity should be avoided.

D. Museum Lighting System:- Light intensity is a primary consideration in museums to protect historic artifacts from damage. 5 to 10 foot-candles (approx. 50 to 100 lux) is currently considered to be the maximum allowable light level for very sensitive materials, such as prints, drawings,watercolors, dyed fabrics, manuscripts, and botanical specimens. Up to 15 footcandles.

E. Open CV: OpenCV is an open source library for image and video analysis, originally introduced more than decade ago by Intel. Since then, a number of programmers have contributed to the most recent library developments. The latest major change took place in 2009 (OpenCV 2) which includes main changes to the C++ interface. Nowadays the library has >;2500 optimized algorithms. It is extensively used around the world, having >;2.5M downloads and >;40K people in the user group. Regardless of whether one is a novice C++ programmer or a professional software developer, unaware of OpenCV, the main library content should be interesting for the graduate students and researchers in image processing and computer vision areas. To master every library element it is necessary to consult many books available on the topic of OpenCV. However, reading such more comprehensive material should be easier after comprehending some basics about OpenCV from this paper.

II. SYSTEM ARCHITECTURE

The system architecture of this proposed system is following.

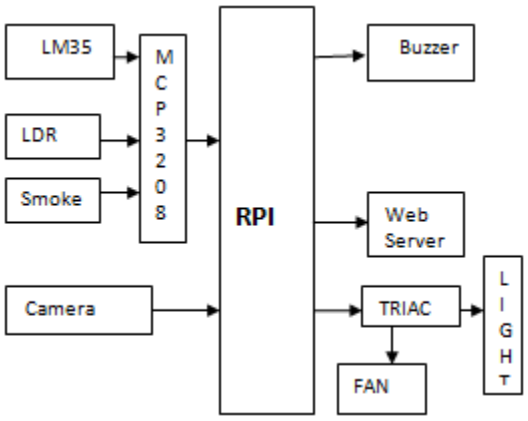


Figure.1. Block Diagram

Raspberry Pi: Hardware implementation for This proposed system is shown in above with the blocks. Raspberry Pi is the processor and its relevant components. The Wi-Fi is used for wireless communication and Wi-Fi USB module is interfaced to Raspberry Pi's USB port and sensor's data is to upload to web server and live monitoring by camera and Buzzer is used for alarm and LCD is used for display the Sensors data and TRIAC is used for switching the fan

and light. When the sensors are data reached threshold limit the buzzer sound will alert.

III. IMPLEMENTATION

A. Hardware

In hardware implementation, Raspberry Pi plays a key role in monitoring in this system. The Raspberry Pi is a small computer, same as the computers with which you're already familiar. It uses a many different kinds of processors, so can't install Microsoft Windows on it. But can install several versions of the Linux operating system that appear and feel very much like Windows. Simple to use but powerful, affordable and in addition difficult to break, Raspberry Pi is the perfect device for aspiring computer scientists. This small computer features amazing HD (high-definition) quality, video playback, also sports high quality audio and has the capability to play 3D games. The device use the ARM processor which does nearly all of the hard work in order to run the Raspberry Pi. The overview of Raspberry Pi has shown below

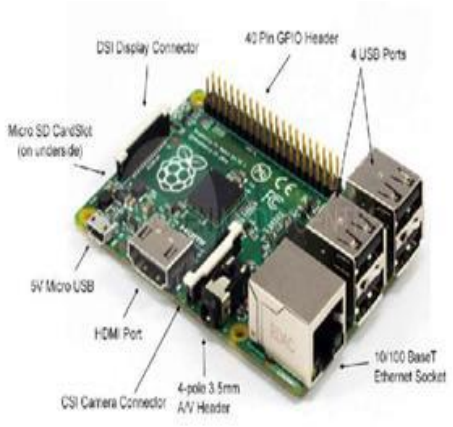


Fig.2. Raspberry Pi

1. GPIO: One powerful feature of the Raspberry Pi is the row of GPIO (general purpose input/output) pins along the Top edge of the board. These pins are physical interface between the pi and the oust side world. At the simplest Level ,you can think of them as switches that you can turn on or off(input) or that the pi can turn on or off(output).Of the 40 pins,26 are GPIO pins and other are power and ground pins. You can program the pins to interact in amazing ways with the real world. Inputs don't have to come from a physical switch. It could be input from a sensor or a signal from another computer or device, for example. The output can also do anything, from turning on LED to sending a signal or data to another device. If the Raspberry Pi is on a network, you can control devices that are attached to it from any where and those devices can send data back. Connectivity and control of physical devices over the internet is a powerful and exciting thing, and Raspberry Pi is ideal for this.

2. Temperature Sensor*:* The temperature sensor will give a variable output voltage with respect to the temperature variation. LM-35 is used as temperature sensor which is a precision integrated-circuit temperature sensor, Calibrated directly in ° Celsius (Centigrade), Linear + 10.0 mV/oC scale factor with accuracy O.soC (at +25°C) with rated for full -55° to +150°C range. The Temperature Sensor which I have used in this project has shown below:

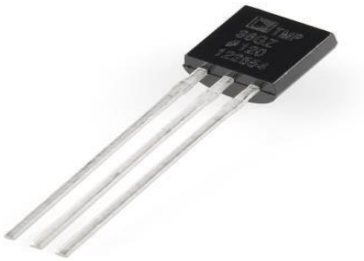


Fig.3. Temperature Sensor

3.Smoke Sensor*:* The smoke sensor will give a variable output voltage with respect to the temperature variation. There is better sensitivity for natural gas and coal gas. The Smoke Sensor which I have used in this project has shown below:

Fig.4. Smoke Sensor

4. LDR: LDR (Light Dependent Resistor) is variable resistor, the resistance of the LDR is inversely proportional to the light intensity, it exhibits maximum resistance in the absence of light and minimum resistance in the presence of light. The LDR which I've used in this project has shown below

Fig.5. LDR

5. MCP3208: MCP3208 devices are successive approximation 12-bit Analog-to-Digital (A/D) Converters with on-board sample and hold circuitry. The MCP3208 is programmable to provide two pseudo-differential input pairs or four single ended inputs. The MCP3208 is programmable to provide four pseudo-differential input pairs or eight single ended inputs. The ADC which I have used in this project has shown below: Differential Nonlinearity (DNL) is specified at ±1 LSB, while Integral Nonlinearity (INL) is offered in ±1 LSB (MCP3208-B) and ±2 LSB (MCP3204/3208-C) versions. Communication with the devices is accomplished using a simple serial interface compatible with the SPI protocol. The devices are capable of conversion rates of up to 100 kbps. The MCP3208 devicesoperate over a broad voltage range (2.7V - 5.5V).

B. CAMERA

.Fig.6. USB Camera

The USB camera Module is interfaced to the Raspberry Pi's USB port. The camera is mainly used to captured the changes in the environment i.e. Motions. The required power supply to operate USB camera will get it from Raspberry Pi only.

1. BUZZER: A buzzer or beeper is an audio signaling device which may be mechanical, electro mechanical, or piezoelectric. Typical uses of buzzers and beepers include devices. The buzzer which have used in this project is shown below fig:

Fig.7. Buzzer

2. TRIAC: The BT136 can be used in circuits of frequency conversion, voltage adjust and control. TRIAC's are widely used in AC power control applications. They are able to switch high voltages and high levels of current, and over both parts of an AC waveform. This makes triac circuits ideal for use in a variety of applications where power switching is needed. One particular use of TRIAC circuits is in light dimmers for domestic lighting, and they are also used in many other power control situations including motor control. The TRAIC which I have used in this project is shown below:

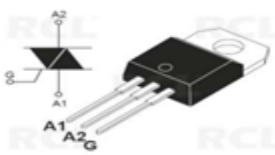


Fig.8. TRIAC.

The TRIAC is a development of the thyristor. While the thyristor can only control current over one half of the cycle, the TRIAC controls it over two halves of an AC waveform. As such the TRIAC can be considered as a pair of parallel but opposite thyristors with the two gates connected together and the anode of one device connected to the cathode of the other, etc. However the names of these are a little more difficult to assign, because the main current carrying terminals are connected to what is effectively a cathode of one thyristor, and the anode of another within the overall device. There is a gate which acts as a trigger to turn the device on. In addition to this the other terminals are both called Anodes, or Main Terminals These are usually designated Anode 1 and Anode 2 or Main Terminal 1 and Main Terminal 2 (MT1 and MT2). When using TRIAC's it is both MT1 and MT2 have very similar properties.

C. Software Here, to program Raspberry Python was used. And a Sever as HTML Web server. Final Schematic Diagram of this Project has shown below:

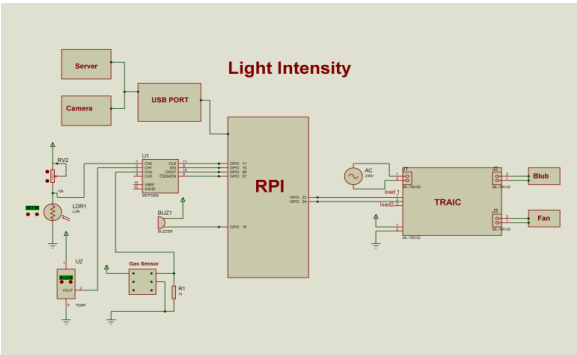


Fig.9. Schematic

IV. ALGORITHM & FLOWCHART

A. Algorithm

- Step– 1: Initialize RPI and camera.
- Step– 2: Taking sensor reading by ADC which have interfaced with RPI .
- Step – 3: uploaded the Sensor's data in to web server and Live streaming by camera which have interfaced with RPI .
- Step– 4:if the sensor's reached the threshold limits the light intensity will vary and fan will turn on and buzzer will turn on.
- Step– 5: continues till system runs.

B. Flowchart

The flowchart of this paper is shown below.

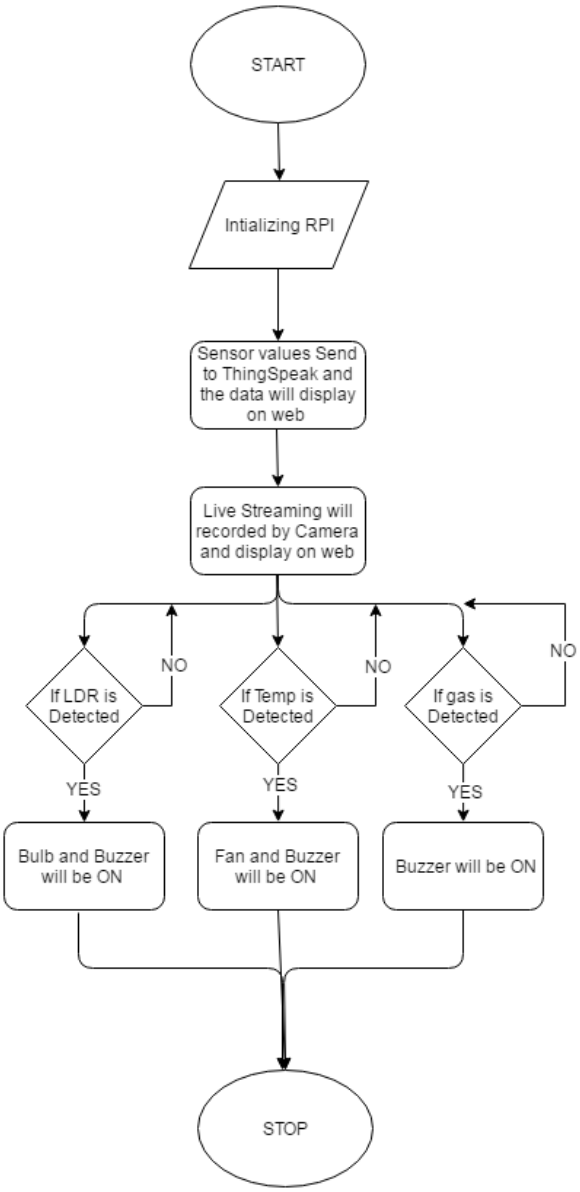


Fig.10. Flow Chart

V. RESULTS

Fig.11. Final Prototype

VI. CONCLUSION

The Facility manger will have skill, training and experience but lagging with lack of information to take action immediately. In the paper, we have proposed and developed cloud based light intensity, temperature and smoke monitoring system. This helps to Facility manger to take necessary action at right time, with proper controlling with can achieve desired results and we can monitor live streaming by camera. To evaluate the system, we have considered laboratory as an example but it can be used at various applications like traffic light monitoring, poultry lighting and museum lighting etc to avoid damages.

VII.REFERENCES

[1]http://www.engineersgarage.com/armprojects /introduction-to-arm-microcontroller-RaspberryPi [2]B.Johnson, “How the Raspberrypi works” Internet: computer.howstuffworks.com/raspberry-pi1.htm
[3]http://www.nex-robotics.com/RaspberryPi development-board/RaspberryPi-development-board.html
[4]https://en.wikipedia.org/wiki/Temperature-sensor [5]https://en.wikipedia.org/wiki/soil - sensor
[6]https://en.wikipedia.org/wiki/Rain fall-sensor

Smart Waste Collection Monitoring and Alert System via IoT

International Conference on Latest Trends in Electronics and Communication (ICLTEC)-ISBN 978-93-88808-62-0

Vaddepally Anusha
M.Tech Scholar
Department of Electronics and Communication Engineering
Mallareddy College of Engineering, Hyderabad, India
anushavaddepally18@gmail.com

***Abstract*—The uncollected waste material when the waste bin is full is a common problem nowadays. Thus, an efficient waste management for the waste material is essential in ensuring a clean and green surrounding environment. This paper presents an Internet of Things (IoT) based Smart Waste Collection Monitoring and Alert System to monitor the waste material at the selected site of garbage collection area. The system is implemented using an ultrasonic sensor which is connected to Arduino UNO as to monitor waste bin garbage level. In this system, waste bin depth level will be sent via Arduino Ethernet Shield with an Internet connection to the Ubidots IoT Cloud. The Ubidots store the collected waste bin level data into IoT database and display the waste bin depth level on online dashboard for real-time visualization. The Ubidots Event manager invoke a notification alert to garbage collector mobile phone via a SMS when the waste bin is nearly filled for immediate waste collection. Therefore, the waste collection became more effective and systematic.**

Keywords— Ultrasonic Sensor, Arduino UNO, Smart Waste Collection Monitoring and Alert System, Ubidots.

I. INTRODUCTION

Currently, over 23,000 tonnes of waste is produced each day in Malaysia. However, this amount is expected to rise to 30,000 tonnes by the year 2020. The amount of waste generated continues to increase due to the increasing population and development, and only less than 5% of the waste is being recycled. Despite the massive amount and complexity of waste produced, the standards of waste management in Malaysia are still poor. Hence, the country will face a problem when there is no proper management for this waste collection. As Malaysia is a developing nation and furthermore has comparative issues, for example, legitimate innovations, labour, arrive shortage and different offices which are deficient to adapt to the regularly expanding rate of waste age[1].

An efficient waste material collection is essential to prevent the waste from affecting human health and polluting environment. Large quantities of uncollected waste material are one of the thing that can affect society health and ecological system if it is not properly managed. The country will face a problem when there is no proper management for this waste collection department. As the country grow, waste material also grows too. Hence, one of the problem to solve the uncollected waste problem by knowing when the waste bin is full and ready for waste collection. This information can help the city council to properly schedule their garbage truck for waste collection within their managed area. This will improve the movement of waste collection fleet resources while enhance the efficiency of waste collection system.

Solid waste management is one of major aspect which has to be considered in terms of making urban area environment healthier. It is become necessary and challenging to manage the solid waste with rapid urbanization and increased population growth. An environment will be polluted and dirty if the waste material is not been manage and collected in time. A better waste management solution can helps improving the general wellbeing of a community and built up a better neighborhood. Nowadays, numerous IoT based solution for waste management are implemented to improve the collection of garbage which would ensure healthy environment for life on this green planet, with greater efficiency [2],[3],[4],[5]. Municipalities wanting to achieve cleaner urban environments can implement an IoT based solution [6],[7]. Some of the IoT based solutions [8],[9],[10] for waste management provide a notification alert when the garbage bin reaches its full capacity for immediate waste collection. An IoT based cost-effective system that can monitor the everyday garbage IoT based solid waste management system which enables garbage bin monitoring, dynamic scheduling and routing of garbage collector trucks in a smart city [11],[12]. A review of existing IoT-enabled solutions in smart cites' waste management is done here to bring together the state-of-the-art solution for example in term of self-powered solution.[13],[14].

IoT which is a new platform that is very useful for people in this world. IoT is the core of such revolutionary of growing engines. IoT is possible due to sufficient power supply and internet connectivity[15]. The term of IoT is commonly used to describe a framework where sensors are connected to objects and help these objects to share their ‘digital voice’ with the external world over internet connection. In the recent time, IoT has become a compilation of purpose-built networks. There were many IoT platforms such as Blynk, Ubidots, IBM Bluemix and Devicepilots as bidots is chosen as a platform of IoT for this project. Ubidots is a cloud service that offers a friendly and intuitive interface where the users can interact with a variety of devices, ranging from a cell phone or a computer, to an embedded system such as a microcontroller system. In a nutshell, Ubidots is a platform that allows to link different types of devices to a cloud database and save variables that can represent them in a simple and fast way and secure manner.

This paper presents a Smart Waste Collection Monitoring and Alert System (SWCMAS) using IoT technology on Ubidots Cloud. The remainder of the paper is organized as follows. Section II describes the system development and overall design approach. Section III discussed the experimental result of an IoT smart waste collection monitoring and alert system performance. Finally, section IV provides the concluding remarks and point out the ideas for future extension of this work.

II. System Development

In this section, a brief explanation regarding on the project development and methodology will be described. This project proposed a system to control a waste material from overflow from the waste bin and alert is send to cleaner for waste collection. Using the anticipated system, monitoring of the waste collection status could be monitored effectively. This project designates a technique in which could monitor the garbage level at regular intervals as overflow of the bin can be prevented. The filling level of the garbage in the dustbin and its original level height could be sensed/ monitored by the ultrasonic sensor. Programming in the Arduino UNO is done in such a way that once a particular level of filling is sensed, information as a message is sent to the user, requesting for cleaning of the dust bin.

Referring to the block diagram in Figure 1 below, the developed system consists of a 1) sensor node that implemented using an Arduino Uno board connected with Arduino Ethernet Shield, HCRS-04 ultrasonic sensor and buzzer; 2) Wired router that interconnected the sensor node to Ubidots IoT Cloud platform and 3) Ubidots Cloud platform that consist secured IoT devices organization.

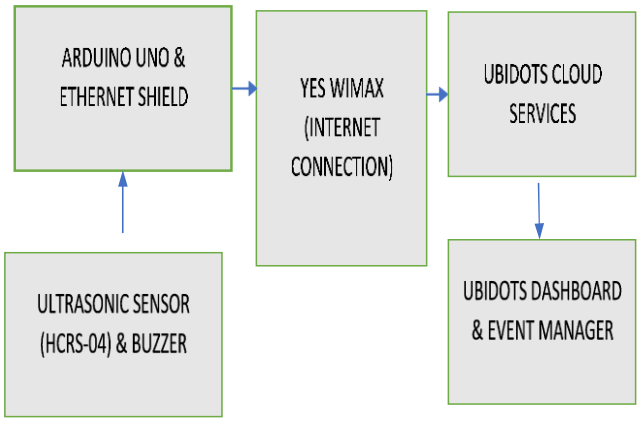


Fig. 1. Block diagram of SWCMAS

Ubidots Dashboard to display and visualize the waste bin depth level data and Ubidots Event manager to notify an alert to consumer via telegram/sms on smartphone when the waste bin is nearly filled.

The system development consists of hardware and software development phase. In hardware development phase, the Arduino Uno is use as a microcontroller for the sensor node. The sensor node is interface with Ultrasonic sensor for detection of the waste bin depth.

A. System Hardware Development

Figure 2 above shows the schematic diagram for connection of Arduino Uno and Ultrasonic sensor. The pin that is used is all digital pins, hence no usage of analog pins. The ground pin of ultrasonic sensor must be connected to ground pin of Arduino Uno to avoid short-circuit. The trig pin and echo pin is connected to (digital) pin 8 and (digital) pin 9 respectively. The RJ45 cable also is connected from Arduino Ethernet Shied to YES WiMAX router for wired internet connection.

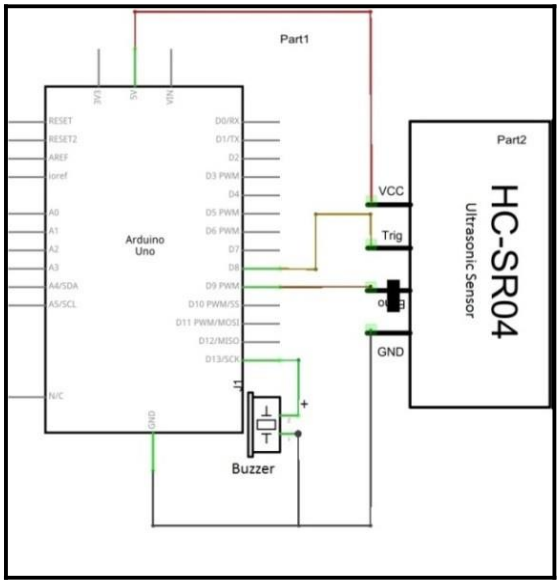


Fig. 2. Schematic diagram of SWCMAS

The ultrasonic sensors are placed at the top of the bin or at the lid of the waste bin as shown in Figure 3. When the waste is started to be filled, the Ultrasonic sensor starts emitting sound waves. At one side is the transmitter and other is the receiver, which measures distance by sending out a sound wave at a specific frequency and listening for that sound wave to bounce back. By recording the elapsed time between the sound wave being generated and the sound wave bouncing back, it is possible to calculate the distance between the ultrasonic sensor and the object. It comes as a whole with an ultrasonic transmitter and collector module. The distance can be calculated with the following formula:

$$\text{Distance } L = 1/2 \times T \times C \quad (1)$$

where L is the distance, T is the time between the emission and reception, and C is the sonic speed. (The value is multiplied by 1/2 because T is the time for go-and-return distance.).

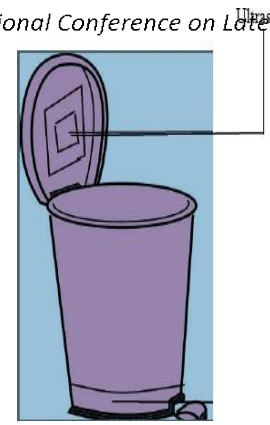


Fig. 3. Waste Bin Model

Figure 4 shown the waste bin at the selected site. The ultrasonic sensors are placed at the top of the bin. The step of the collection data will be repeated as using the waste bin model. The data or information is collected and send to Ubidots platform.

Fig. 4. Waste Bin at Selected Site

B. Software Development and Configuration

In software development and configuration phase, the MAC and IP address of sensor node have to be obtained and configure into Arduino program for sending the waste bin depth level to Ubidots IoT Cloud.

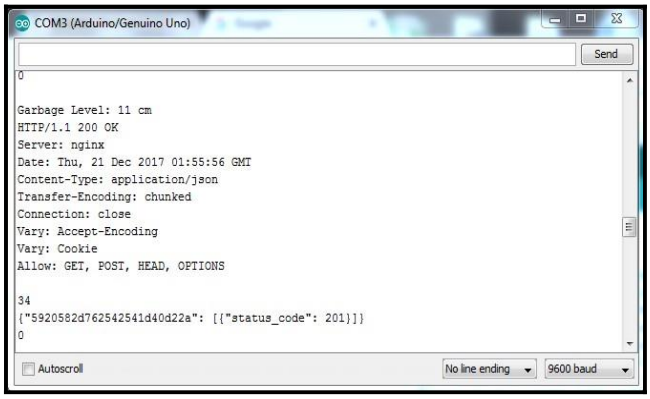


Fig. 5. Ultrasonic sensor data publish to Ubidots Cloud

Figure 5 shows the the waste bin depth level data collection from the sensor node and publish the collected data using MQTT protocol connection to Ubidots platform in Arduino IDE. Figure 5 also shows that the transfer-encoding for this project is chunked and vary in accept-encoding and cookie. The connection is established as the connection from the Arduino IDE and Ubidots website is close.

After many steps of microcontroller coding using Arduino, Node configuration, Arduino Ethernet Shield configuration, the data is finally collected and sent to Ubidot IoT Cloud. Then, the project proceeded by configuring Ubidots platform IoT organization. The interface of Ubidots is user-friendly and ensures quick visualization of the multiples of interest.

By using this platform, you can create an event or warning messages, send them to a mobile device and configure the device to execute a control action. The system will alert the user and will send a message to user through cell phone via Wi-Fi. The device must be built in the Ubidots platform to take the devices API ID as to synchronize with Arduino IDE coding. Within Ubidots, the data can be retrieved in an instant. The event will be created as shown as Figure 6 below to notify the user when the waste value is less than 4 cm.

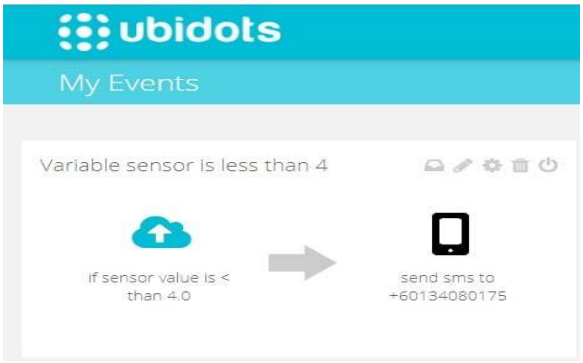


Fig. 6. The Event Setup for SMS notification Alert

III. RESULT

This section shown the results of data gathered during this project implementation. All of the data and information details were collected to evaluate the system performance. The result consists of two parts. The first part is an Ultrasonic sensor distance result that indicate the depth level of waste bin data display in the serial monitor on Arduino IDE. The second part is obtained from depth level of waste bin data collection through IoT Cloud on Ubidots Dashboard which is an online display. By creating a device in Dashboard, we can collect and store the waste bin depth level data on the specified device. The notification alert can be invoked when certain value of device data changes more or less from the specified value.

A. Arduino IDE results

Figure 7 shows the result on the serial monitor on Arduino IDE that indicate the depth level of waste bin data.

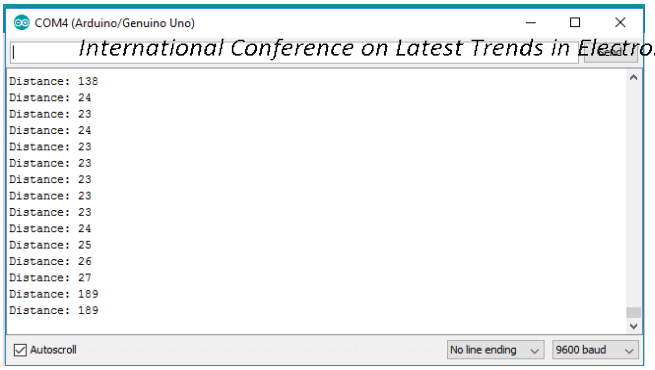


Fig. 7. The Ultrasonic Sensor Distance Measurement Result

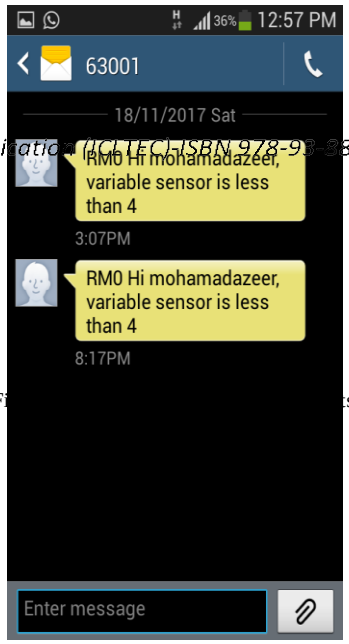
B. *Ubidots Dashboard results*

The performances of this waste collection monitoring system can be monitored through Ubidots Dashboard. The data or information of the distance value can be view in raw, average or overall. The data and information that is collected is from waste bin from the selected site. The data will be sends from Sensor node through Arduino Ethernet Shield to Ubidots devices that had been created. Figure 8 shows the graph of the collected data of distance between garbage level and the waste bin lid while Figure 9 shows the value of the depth of the garbage level to the lid chronological order as the waste bin is nearly full.

Fig. 8. Waste bin depth level data graph at Ubidots

2017-11-18 20:16:54 +08:00	2	
2017-11-18 20:16:26 +08:00	7	
2017-11-18 20:16:24 +08:00	11	
2017-11-18 20:16:22 +08:00	12	
2017-11-18 20:16:20 +08:00	23	
2017-11-18 20:16:16 +08:00	33	
2017-11-18 20:16:14 +08:00	31	
2017-11-18 20:16:10 +08:00	25	
2017-11-18 20:16:08 +08:00	23	
2017-11-18 20:16:05 +08:00	17	
2017-11-18 20:15:57 +08:00	11	
2017-11-18 20:15:27 +08:00	60	
2017-11-18 20:15:21 +08:00	63	
2017-11-18 20:15:18 +08:00	373	

Fig. 9. The Waste bin depth level on Ubidots Cloud



Fi ts.

As the garbage level increase, the distance between the waste bin lid with ultrasonic sensor will decrease Once the level of distance less than the determined value (4 cm), Ubidots will alert and send message to authorized person in form of SMS as shown in Figure 10.

IV. Conclusion

In this paper, we propose a new solution to enhance waste collection efficiently using the Arduino Uno with Arduino Ethernet Shield technology and ultrasonic sensor systems. In this proposed system, the garbage overflow of garbage can be avoided and managed efficiently. This will intimate or send SMS or email to the authorized person through Ubidots platform. The garbage managing system and the facility of collecting the garbage presently doesn't fit to the current requirement. Hence better facility of collecting garbage and transportation should be provided. Since, this system provides the information when the bin gets completely filled with garbage, it reduces the number of times the arrival of vehicle which collects the garbage. This method finally helps in keeping the environment clean. Thus, the waste collection is made more efficient.

Acknowledgment

We would like to thanks the sponsors of development of Ubidots for free evaluation license on Industrial IoT and also provide us with the reading materials supported hardware libraries.

REFERENCES

[1] J. Sreenivasan, M. Govindan, M. Chinnasami, and I. Kadiresu, "Solid Waste Management in Malaysia – A Move Towards Sustainability," *Waste Manag. An Integr. Visions*, vol. 2005, no. April 2005, pp. 55–70, 2012.

[2] M. Adam, M. E. Okasha, O. M. Tawfeeq, M. A. Margan and B. Nasreldeen, "Waste Management System Using IoT," 2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), Khartoum, 2018, pp. 1-4

[3] S. Aleyadeh and A. M. Taha, "An IoT-Based Architecture for Waste Management," 2018 IEEE International Conference on Communications Workshops (ICC Workshops), Kansas City, MO, 2018, pp. 1-4.

[4] G. K. Shyam, S. S. Manvi and P. Bharti, "Smart waste management using Internet-of-Things (IoT)," 2017 2nd International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 2017, pp. 199-203.

[5] W. Chen, Y. Wang, P. Huang, Y. Huang and M. Tsai, "A Smart IoT System for Waste Management," *2018 1st International Cognitive Cities Conference (IC3)*, Okinawa, 2018, pp. 202-203.

[6] S. S. Chaudhari and V. Y. Bhole, "Solid Waste Collection as a Service using IoT-Solution for Smart Cities," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, 2018, pp. 1-5.

[7] E. Al-Masri, I. Diabate, R. Jain, M. H. Lam and S. Reddy Nathala, "Recycle.io: An IoT-Enabled Framework for Urban Waste Management," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 5285-5287

[8] S. Paul, S. Banerjee and S. Biswas, "Smart Garbage Monitoring Using IoT," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 1181-1185.

[9] H. N. Saha, S. Gon, A. Nayak, S. kundu and S. Moitra, "Iot Based Garbage Monitoring and Clearance Alert System," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 204-208.

[10] S. K. Memon, F. Karim Shaikh, N. A. Mahoto and A. Aziz Memon, "IoT based smart garbage monitoring & collection system using

[11] G. K. Shyam, S. S. Manvi and P. Bharti, "Smart waste management using Internet-of-Things (IoT)," 2017 2nd International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 2017, pp. 199-203.

[12] W. Chen, Y. Wang, P. Huang, Y. Huang and M. Tsai, "A Smart IoT System for Waste Management," *2018 1st International Cognitive Cities Conference (IC3)*, Okinawa, 2018, pp. 202-203.

[13] S. S. Chaudhari and V. Y. Bhole, "Solid Waste Collection as a Service using IoT-Solution for Smart Cities," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, 2018, pp. 1-5.

[14] E. Al-Masri, I. Diabate, R. Jain, M. H. Lam and S. Reddy Nathala, "Recycle.io: An IoT-Enabled Framework for Urban Waste Management," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 5285-5287

[15] S. Paul, S. Banerjee and S. Biswas, "Smart Garbage Monitoring Using IoT," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 1181-1185.

[16] G. K. Shyam, S. S. Manvi and P. Bharti, "Smart waste management using Internet-of-Things (IoT)," 2017 2nd International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 2017, pp. 199-203.

[17] W. Chen, Y. Wang, P. Huang, Y. Huang and M. Tsai, "A Smart IoT System for Waste Management," *2018 1st International Cognitive Cities Conference (IC3)*, Okinawa, 2018, pp. 202-203.

[18] S. S. Chaudhari and V. Y. Bhole, "Solid Waste Collection as a Service using IoT-Solution for Smart Cities," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, 2018, pp. 1-5.

[19] E. Al-Masri, I. Diabate, R. Jain, M. H. Lam and S. Reddy Nathala, "Recycle.io: An IoT-Enabled Framework for Urban Waste Management," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 5285-5287

[20] S. Paul, S. Banerjee and S. Biswas, "Smart Garbage Monitoring Using IoT," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 1181-1185.

[21] H. N. Saha, S. Gon, A. Nayak, S. kundu and S. Moitra, "Iot Based Garbage Monitoring and Clearance Alert System," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 204-208.

[22] G. K. Shyam, S. S. Manvi and P. Bharti, "Smart waste management using Internet-of-Things (IoT)," 2017 2nd International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 2017, pp. 199-203.

[23] W. Chen, Y. Wang, P. Huang, Y. Huang and M. Tsai, "A Smart IoT System for Waste Management," *2018 1st International Cognitive Cities Conference (IC3)*, Okinawa, 2018, pp. 202-203.

[24] S. S. Chaudhari and V. Y. Bhole, "Solid Waste Collection as a Service using IoT-Solution for Smart Cities," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, 2018, pp. 1-5.

[25] E. Al-Masri, I. Diabate, R. Jain, M. H. Lam and S. Reddy Nathala, "Recycle.io: An IoT-Enabled Framework for Urban Waste Management," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 5285-5287

[26] S. Paul, S. Banerjee and S. Biswas, "Smart Garbage Monitoring Using IoT," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 1181-1185.

[27] H. N. Saha, S. Gon, A. Nayak, S. kundu and S. Moitra, "Iot Based Garbage Monitoring and Clearance Alert System," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 204-208.

[28] G. K. Shyam, S. S. Manvi and P. Bharti, "Smart waste management using Internet-of-Things (IoT)," 2017 2nd International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 2017, pp. 199-203.

[29] W. Chen, Y. Wang, P. Huang, Y. Huang and M. Tsai, "A Smart IoT System for Waste Management," *2018 1st International Cognitive Cities Conference (IC3)*, Okinawa, 2018, pp. 202-203.

[30] S. S. Chaudhari and V. Y. Bhole, "Solid Waste Collection as a Service using IoT-Solution for Smart Cities," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, 2018, pp. 1-5.

[31] E. Al-Masri, I. Diabate, R. Jain, M. H. Lam and S. Reddy Nathala, "Recycle.io: An IoT-Enabled Framework for Urban Waste Management," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 5285-5287

[32] S. Paul, S. Banerjee and S. Biswas, "Smart Garbage Monitoring Using IoT," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 1181-1185.

[33] H. N. Saha, S. Gon, A. Nayak, S. kundu and S. Moitra, "Iot Based Garbage Monitoring and Clearance Alert System," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 204-208.

[34] G. K. Shyam, S. S. Manvi and P. Bharti, "Smart waste management using Internet-of-Things (IoT)," 2017 2nd International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 2017, pp. 199-203.

[35] W. Chen, Y. Wang, P. Huang, Y. Huang and M. Tsai, "A Smart IoT System for Waste Management," *2018 1st International Cognitive Cities Conference (IC3)*, Okinawa, 2018, pp. 202-203.

[36] S. S. Chaudhari and V. Y. Bhole, "Solid Waste Collection as a Service using IoT-Solution for Smart Cities," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, 2018, pp. 1-5.

[37] E. Al-Masri, I. Diabate, R. Jain, M. H. Lam and S. Reddy Nathala, "Recycle.io: An IoT-Enabled Framework for Urban Waste Management," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 5285-5287

[38] S. Paul, S. Banerjee and S. Biswas, "Smart Garbage Monitoring Using IoT," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 1181-1185.

[39] G. K. Shyam, S. S. Manvi and P. Bharti, "Smart waste management using Internet-of-Things(IoT)," 2017 2nd International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 2017, pp. 199-203.

[40] W. Chen, Y. Wang, P. Huang, Y. Huang and M. Tsai, "A Smart IoT System for Waste Management," *2018 1st International Cognitive Cities Conference (IC3)*, Okinawa, 2018, pp. 202-203.

[41] S. S. Chaudhari and V. Y. Bhole, "Solid Waste Collection as a Service using IoT-Solution for Smart Cities," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, 2018, pp. 1-5.

[42] E. Al-Masri, I. Diabate, R. Jain, M. H. Lam and S. Reddy Nathala, "Recycle.io: An IoT-Enabled Framework for Urban Waste Management," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 5285-5287

[43] S. Paul, S. Banerjee and S. Biswas, "Smart Garbage Monitoring Using IoT," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 1181-1185.

[44] H. N. Saha, S. Gon, A. Nayak, S. kundu and S. Moitra, "Iot Based Garbage Monitoring and Clearance Alert System," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 204-208.

[45] S. K. Memon, F. Karim Shaikh, N. A. Mahoto and A. Aziz Memon, "IoT based smart garbage monitoring & collection system using WeMos & Ultrasonic sensors," 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, 2019, pp. 1-6.

[46] R. N. Subrahmanian, S. K. S, T. Suvidharbabu, and B. Manikandan, "Smart Bin using IoT with Fog Computing," International Journal of Emerging Technology in Computer Science & Electronics vol. 24, no. 6, pp. 13–16, 2017.

[47] S. Murugaanandam, V. Ganapathy and R. Balaji, "Efficient IOT Based Smart Bin for Clean Environment," 2018 International Conference on Communication and Signal Processing (ICCSP), Chennai, 2018, pp. 0715-0720.

[48] N. Alsbou, M. A. Samad, M. Alhashem and A. S. A. Abuabed, "Developing a Self-Powered Enlarging Smart Waste Bin," 2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC), Limassol, 2018, pp. 683-689.

[49] S. Mdukaza, B. Isong, N. Dladlu and A. M. Abu-Mahfouz, "Analysis of IoT-Enabled Solutions in Smart Waste Management," IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society, Washington, DC, 2018, pp. 4639-4644.

[50] D. Evans, "The Internet of Things How the Next Evolution of the Internet" Cisco Internet Business Solutions Group (IBSG) no. April, 2011. http://www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf

Chip Design for Turbo Encoder Module for In-Vehicle System

PULUKURI NEIL SAMUEL SADANAND,MTECH(ES&VLSI),MRCE

***Abstract*—This paper studies design and implementation of the Turbo encoder to be an embedded module in the in-vehicle system (IVS) chip. Field programmable gate array (FPGA) is employed to develop the Turbo encoder module. Both serial and parallel computations for the encoding technique are studied. The two design methods are presented and analyzed. Developing the parallel computation method, it is shown that both chip size and processing time are improved. The logic utilization is enhanced by 73% and the processing time is reduced by 58%. The Turbo encoder module is designed, simulated, and synthesized using Xilinx tools. Xilinx Zynq-7000 is employed as an FPGA device to implement the developed module. The Turbo encoder module is designed to be a part of the IVS chip on a single programmable device.**

***Index Term:* Turbo encoder module; field programmable gate array; emergency call; in-vehicle system chip.**

I. Introduction

The European emergency call (eCall) system is a telematics system designed to save more lives in vehicle accidents. It is a governmental mandatory system that is to be implemented by March 2018 [1][2]. The EU eCall system provides an immediate voice and data channel between the vehicles and an emergency center after car accidents. The data channel provides the emergency center with the necessary data for emergency aids.

The EU eCall system main parts includes the in-vehicle system (IVS), the public safety answering point (PSAP), a cellular communication channel. The IVS activates the data channel automatically when a car accident occurs. The IVS collects the minimum set of data (MSD) that includes GPS coordinates, the VIN number, and all required data for an emergency aid. It sends the MSD to the closest PSAP through a cellular channel in up to 4 seconds [1]. The PSAP sends the emergency team to the location of the accidents.

The IVS modem employs multiple modules for the MSD signal processing. The modules of the IVS are shown in Figure 1. The IVS employs a Turbo encoder as a forward error correcting (FEC) [1]. The Turbo encoder implements the digital data encoding technique in data transmissions. Turbo coding is one of the most popular and efficient coding technique to improve bit error rate (BER) in digital communications [3] [4]. The cyclic redundancy check (CRC) [5], the modulator [6], the demodulator-decoder [7] modules are projected and implemented on an FPGA device. They are developed to be embedded modules of the IVS chip.

978-1-5386-4649-6/18/$31.00

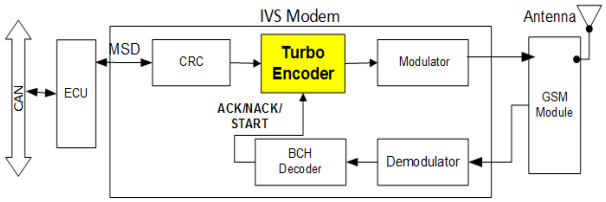


Fig. 1: The IVS block diagram.

This work studies the hardware development of the Turbo encoder. It employs FPGA technologies to develop the Turbo encoder to be an embedded module in the IVS modem. It discusses serial and parallel computation techniques for the Turbo encoder. It does not only design and implement the Turbo encoder module, but also proposes a better solution for the turbo encoder implementation. The improvement of the chip size and processing time are exhibited by developing the parallel computation technique for the Turbo encoder.

II. Turbo Encoder Module

The turbo encoder technique is one of the most powerful FEC techniques in digital communication [8]. The IVS employs a Turbo encoder module with $1/3$ code rate. The Turbo encoder functionalities are detailed in the third generation partnership project (3GPP) standards. The 3GPP Turbo encoder is illustrated in Figure 2 [8]. The input signal of the turbo encodes is the MSD data appended with the CRC parity bits in binary. The block length of the MSD data is 1148 bits. The output of the module is the MSD encoded data in binary. Implementing the turbo coding technique with $1/3$ coding rate and thrills bits, the length of the output is 3456 bits. The thrills structure has an impact of the Turbo encoder [9].

The Turbo encoder employs a parallel concatenated convolutional code (PCCC). The PCCC uses two constituent encoders with eight states as it is shown in Figure 2. The initial status of the register are zeros. The first constituent takes the MSD bits and implements the employed convolutional technique. It takes one bit at a time and generates one bit of

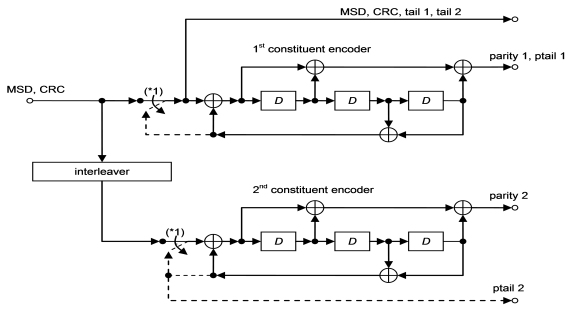


Fig. 2: The structure of the Turbo encoder.

parity1 bits. The second constituent implements an identical technique of the first constituent, but it calls for the MSD bit after they are interleaved with a 3GPP designed interleaver technique [8].

The length of the input data, parity1, and parity2 are 1148 bits. There are 12 bits of the tail bits. They are driven from the shift register feedback. The tail bits are applied for end points between the encoded data blocks. The output structure of the Turbo encoder is illustrated in Figure 3.

MSD+CRC	$tail_1$	$tail_2$	Parity 1	$ptail_1$	Parity 2	$ptail_2$

Fig. 3: The output buffer of the Turbo encoder.

Interleaver

Denote the transfer function of the employed PCCC as:

$$G(D) = \left(1, \frac{g_1(D)}{g_0(D)}\right) \tag{1}$$

where

$$g_(D) = 1 + D^2 + D^3$$

$$g_(D) = 1 + D + D^3$$

And denote the input bits to the encoder as $x_1, x_2, ..., x_K$, the output of the interleaver as $x'_1, x'_2, ..., x'_K$, and the output bits of the first and second constituents as $z_1, z_2, ..., z_K$ and $z'_1, z'_2, ..., z'_K$, respectively; where K is the number of input bits to the Turbo encoder.

The encoder output is expressed as:

$$d_K^{(0)} = x_K, d_K^{(1)} = z_K, d_K^{(2)} = z'_K$$

where $K = 0, 1, ..., K-1$.

The three code blocks of the output, $d_K^{(0)}$, $d_K^{(1)}$, and $d_K^{(2)}$ are separated by trellis bits. The trellis bits are generated from the tail bits of the shift registers after encoding of all the input bits. In figure 2, when the upper switch is lowered and the second constituent is disabled, the three tail bits are used to terminate the first constituent. The output bits of the Turbo encoder, including the trellis bits can be expressed as:

$$d_K^{(0)} = x_K, \ d_{K+1}^{(0)} = z_{K+1}, \ d_{K+2}^{(0)} = x'_K, \ d_{K+3}^{(0)} = z'_{K+1}$$

$$d_K^{(1)} = z_K, \ d_{K+1}^{(1)} = x_{K+2}, \ d_{K+2}^{(1)} = z'_K, \ d_{K+3}^{(1)} = x'_{K+2}$$

$$d_K^{(2)} = x_{K+1}, d_{K+1}^{(2)} = z_{K+2}, d_{K+2}^{(2)} = x'_{K+1}, d_{K+3}^{(2)} = z'_{K+2}$$

where $K = 0, 1, ..., K-1$.

The internal interleaver of the 3GPP Turbo encoder is designed to generate a systematic relationship between x_K and x'_K for any $40 \leq K \leq 5114$ [8]. There is a specific approach to design an internal interleaver for the employed Turbo encoder that is detailed in [8]. This work employs the 3GPP standard approach to design the internal interleaver for the employed Turbo encoder.

First, the input bits of the Turbo encoder is re-arranged in a matrix form that has column, C, and row, R. The rows are labeled as $0, 1, ..., R-1$ and the columns are organized as $0, 1, ..., C-1$. The numbers of rows and columns are determined according to the 3GPP standard for Turbo encoder interleavers [8].

Then the input bits $x_1, x_2, ..., x_K$ are re-organized in a matrix where $y_k = x_k$ for $k = 1, 2, ..., K$, and $y_k = 0$ for the elements that $R \times C > K$:

$$\begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_C \\ y_{(C+1)} & y_{(C+2)} & y_{(C+3)} & \cdots & y_{(C+C)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{((R-1)C+1)} & y_{((R-1)C+2)} & y_{((R-1)C+3)} & \cdots & y_{(R\times C)} \end{bmatrix}$$

Then an intra-row and inter-row permutation is performed on the $R \times C$ matrix. This work employs the 3GPP standard approaches for the intra-row and inter-row.

Denote,

$$s(j) = (v \times s(j-1)) \mod p \tag{2}$$

where $\left\langle s(j) \right\rangle$ for $j \in 1, 2, ..., p-2$ and $p(0) = 0$ is the inta-row permutation sequence and v is associated primitive root for the specified p from table **??**, and use table I to choose the appropriate pattern to compute the inter-row permutation, $\left\langle T(i) \right\rangle$ for $i \in 0, 1, ..., R-1$. i is the row number index of $R \times C$ matrix, and j is the column number index of the matrix.

Also the minimum prime integer (q_i) is determined in the sequence $\left\langle q(i) \right\rangle$ for $i \in 0, 1, ..., R-1$ such that $q_i > q_(i-1)$, $q_i > 6$ and g.c.d$(q_i, p-1) = 1$, where g.c.d is the greater common divisor.

Then one can build a sequence of the permuted prime integers $\left\langle r(i) \right\rangle$ for $i \in 0, 1, ..., R-1$ such that,

$$r_T(i) = q_i, i = 0, 1, ..., R-1$$

TABLE I: 3GPP inter-row permutation pattern

K	*R*	Inter-row permutation patterns <*T*(0), *T*(1), ..., *T*(*R* - 1)>
($40 \leq K \leq 159$)	5	<4, 3, 2, 1, 0>
($160 \leq K \leq 200$) or ($481 \leq K \leq 530$)	10	<9, 8, 7, 6, 5, 4, 3, 2, 1, 0>
($2281 \leq K \leq 2480$) or ($3161 \leq K \leq 3210$)	20	<19, 9, 14, 4, 0, 2, 5, 7, 12, 18, 16, 13, 17, 15, 3, 1, 6, 11, 8, 10>
K = any other value	20	<19, 9, 14, 4, 0, 2, 5, 7, 12, 18, 10, 8, 13, 17, 3, 1, 16, 6, 15, 11>

Denote the pattern of $i-th$ row Intra-row permutation as,

$\left\langle U_i(j) \right\rangle$ for $i \in 0, 1, ..., R-1$,
one can perform the intra-row permutation such that the position of the $i-th$ permuted bit of the $j-th$ row ($U_i(j)$) is calculated based on equation 3, 4, or 5.

if $(C = p)$ then,

$$U_i(j) = s((j \times r_i) \mod (p-1)) \tag{3}$$

where $j = 0, 1, ..., (p-1)$ and $U_i(p-1) = 0$.

if $(C = p+1)$ then,

$$U_i(j) = s((j \times r_i) \mod (p-1)) \tag{4}$$

where $j = 0, 1, ..., (p-1)$, $U_i(p-1) = 0$, and $U_i(p) = p$.

if $(C = p-1)$ then,

$$U_i(j) = s((j \times r_i) \mod (p-1)) - 1 \tag{5}$$

where $j = 0, 1, ..., (p-1)$.

And then the inter-row permutation is implemented on the $R \times C$ matrix by using the sequence pattern $\left\langle T(i) \right\rangle$ for $i \in 0, 1, ..., R-1$.

After the permutations, the elements of the $R \times C$ matrix is denoted by $y'_k = y_k$ such that:

$$\begin{bmatrix} y'_1 & y'_{(R+1)} & y'_{(2R+1)} & \cdots & y'_{((C-1)R+1)} \\ y'_2 & y'_{(R+1)} & y'_{(2R+2)} & \cdots & y'_{((C-1)R+2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y'_R & y'_{(2R)} & y'_{(3R)} & \cdots & y'_{(R \times C)} \end{bmatrix}$$

The output of the interleaver, $x'_2, x'_2, ..., x'_K$, is the bit sequence that are read out from the $R \times C$ matrix column by column starting with y'_1 and ending with $y'_{(R \times C)}$. The appended zero bits for $R \times C > K$ are removed from the output.

There are 1148 elements in the interleaver matrix. The 1148 bits of the MSD data are the elements of the interleaver matrix. The interleaver reorganizes the MSD bits in a systematic order. The intra-row and inter-row techniques of the interleaver elements organizations.

Denote the MSD bits as $B_1, B_2, ..., B_K$, where $K = 1148$. According to the algorithm that is explained in the above mathematical modeling, the interleaver matrix is a rectangular matrix, where R is the number of rows and C is the number of columns of the matrix. The interleaver matrix is designed for an input data block that consists of 1148 bits. As a result, the size of the matrix is 20 X 58. The following steps are implemented to drive the interleaver matrix:

1. The input bits of the matrix are denoted as $b_1, b_2, ..., b_K$, where $b_K = B_K$ and $K = 1148$. The remaining elements are padded with zeros.

2. The intra-row and inter-row permutation are performed according to 3GPP.

3. The calculated elements of the interleaver matrix, except for the padded bits, are stored in a file in hexadecimal format.

The Turbo encoder is developed in Verilog HDL language. Verilog has ability to read the hexadecimal file to get the data and use it as the interleaver data.

The length of the input data, parity1, and parity2 are 1148 bits. There are 12 bits of the tail bits. They are driven from the shift register feedback. The tail bits are applied for end points between the encoded data blocks. The output structure of the Turbo encoder is illustrated in Figure 3.

The employed interleaver for the Turbo encoder is designed according to 3GPP standards [8]. The interleaver elements are organized in a rectangular matrix. There are 1148 elements in the interleaver matrix. The 1148 bits of the MSD data are the elements of the interleaver matrix. The interleaver reorganizes the MSD bits in a systematic order. There are intra-row and inter-row techniques of the interleaver elements organizations.

Denote the MSD bits as $B_1, B_2, ..., B_K$, where $K = 1148$. According to the 3GPP standards [8], the interleaver matrix is a RxC rectangular matrix, where R is the number of rows and C is the number of columns. The size of the matrix is 20 X 58. The following steps are implemented to drive the interleaver matrix:

1. The input bits of the matrix are denoted as $b_1, b_2, ..., b_K$, where $b_K = B_K$ and $K = 1148$. The remaining elements are padded with zeros.

2. The intra-row and inter-row permutation is performed according to 3GPP [8].

3. The calculated elements of the interleaver matrix, except for the padded bits, are stored in in a file in hexadecimal format.

The Turbo encoder is developed in Verilog HDL language. Verilog has ability to read the hexadecimal file to get the data and use it as the interleaver data.

III. FPGA Design for the Turbo Encoder Module

FPGA technologies are employed to develop and implement the designed Turbo encoder module. The register transfer level (RTL) of the module is developed in Verilog HDL. There are multiple registers defined for the input, output, and necessary parameters to implement the Turbo encoding technique. This work studies two methods to execute the encoding, which are serial computation and parallel computation.

The serial computation method processes one bit in one clock cycle. It reads the input data of the MSD, builds the input and output registers, and calculates the parity1, parity2, and the tail bits in a serial process. After performing the encoding, it generates the output bits. Although the method is designed and implemented, it is noted that there is a long processing time that can be overlapped with the other processes in the module. Figure 4 shows the pseudocode of the serial computation of the Turbo encoder module.

The parallel computing technique is employed to develop the Turbo encoder in Verilog. There are many processes in the serial computation technique that are overlapped by using parallel computing technique. There are two functions developed in the parallel Turbo encoder. The two functions implements almost all the processing time of the encoding technique. The Turbo encoding technique needs the MSD data as a whole package to implement the encoding. Figure 5 shows the pseudocode of the implemented parallel technique

```
Pseudocode code for Serial Computation of Turbo encoder
module TURBO_SERIAL ( inputs, outputs;)
        define REGISERS and PARAMETERS;
      always @(posedge clock, posedge reset)
        begin
         if (reset) output=0;
         else begin
            repeat1 (1148) {
                      Read MSD input data}
             If (repeat1 is done)
                repeat2 (1148) {
                      Build output register for MSDinput;
                      Build output register for Parity1;
                      Build output register for Parity2; }
             If (repeat2 is done)
                repeat3 (1148) {
                      Process Parity1 bits; }
             If (repeat3 is done)
                Repeat4 (3) {
                      Process tail1 bits;
                      Process ptail1 bits;    }
             If (repeat4 is done)
                Repeat5 (1148) {
                      Process Parity2 bits; }
             If (repeat5 is done)
                Repeat6(3) {
                      Process tail2 bits;
                      Process ptail2 bits;    }
             If (repeat6 is done)
                Repeat7(3456) {
                      Generate output bits    }
              end  end
endmodule;
```

Fig. 4: The pseudocode for serial computation of the Turbo encoder.

for the Turbo encoder. Both Pseudocodes of the serial and parallel computing techniques are designed in Verilog and implemented on an FPGA device.

The processing time of the Turbo encoder module (in clock cycles) is denoted by T_s for the serial computation and by T_p for the parallel technique, one has,

$$T_s = T_r + T_b + T_{parity1} + T_{tail1} + T_{parity2} + T_{tail2} + T_w \quad (6)$$

$$T_s = 1148 + 1148 + 1148 + 3 + 1148 + 3 + 3456 = 8054$$

where T_r is the time for reading the 1148 bits of the MSD, T_b is the processing time to build the output register, T_{parit1} is the time of processing parity bits, T_{tail} is the time of processing tail bits, and T_w is the time of generating output bits.

Note that the $T_r + T_b + T_{parity1} + T_{tail1} + T_{parity2} + T_{tail2}$ are processed in one clock cycle in the parallel computing technique.

$$T_p = 1 + T_w = 1 + 3456 = 3457 \quad (7)$$

Then one has,

$$T_p = 0.42 T_s \quad (8)$$

Eq. 8 reveals that the parallel computing can improve the proceeding time of the Turbo encoder by 58%.

A. Simulation and Verification

Xilinx tools are utilized to simulate the developed modules. A test bench is designed to simulate the Turbo encoder. Verilog

```
Pseudocode code for Parallel Computation of Turbo encoder
module TURBO_PARALLEL ( inputs, outputs;)
        define REGISERS and PARAMETERS;
 function [3455:0]  codedMSD1;
              input [1147:0] MSDdata;
                 begin
                    repeat1 (1148) {
                       Build output register for MSDinput;
                       Build output register for Parity1;
                       Build output register for Parity2; }
                   If (repeat1 is done)
                      repeat2 (1148) {
                       call  CodedMSD2 (codedMSD1) }
                   end
       endfunction
 function [3455:0]  codedMSD2;
              input [3455:0] codedMSD1;
                 begin
                 repeat3 (1148) {
                       Process Parity1 bits; }
                  If (repeat3 is done)
                    repeat4 (3) {
                       Process tail1 bits;
                       Process ptail1 bits;    }
                If (repeat4 is done)
                    repeat5 (1148) {
                       Process Parity2 bits; }
                If (repeat5 is done)
                    repeat6(3) {
                       Process tail2 bits;
                       Process ptail2 bits;    }
                   end
       endfunction
   always @(posedge clock, posedge reset)
       begin
        if (reset) output=0;
        else begin
           repeat1 (1148) {
                    Read MSD input data    }
              If (repeat1 is done) {
                  call  CodedMSD1 (MSDdata)   }
              Repeat7 (3456) {
                    Generate output bits    }
            end  end
endmodule;
```

Fig. 5: The pseudocode for parallel computation of the Turbo encoder.

HDL is employed to design the test bench. There are two MSD data that are simulated for each of the serial computation and parallel computation of the Turbo encoder. Figure 6 shows the simulation results for the serial computation of the Turbo encoder module. The simulation indicates that the structure output data is correct. However, there is a long time that can be overlapped, which is colored in red in the output trace.

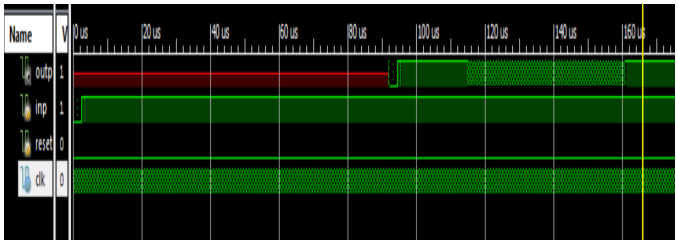


Fig. 6: The simulation result of the Turbo encoder with serial computation.

Figure 7 shows the simulation result of the parallel computing technique. The module starts to generate the output in the beginning. Note that the MSD is encoded in a shorter time compare with the serial computation simulation. This is the

result of parallel computing of multiple process in one clock cycle.

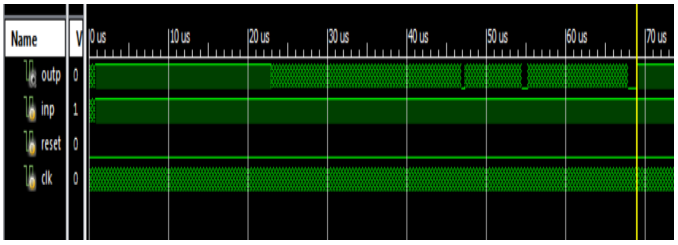


Fig. 7: The simulation result of the Turbo encoder with parallel computation.

B. Hardware Implementation

FPGA has been widely used in the SoC design [10][11]. Zynq-7000 FPGA device is employed to implement the developed Turbo encoder module. Both serial and parallel computation methods are implemented on the FPGA device separately. The Turbo encoder module is synthesized and loaded on the FPGA device. The synthesis reports show that the parallel computation does not only save processing time, but also reduces the utilized logics and the chip size. Figures 8 and 9 show the utilization flip flops and look-up tables (LTU) for the serial and parallel computation respectively.

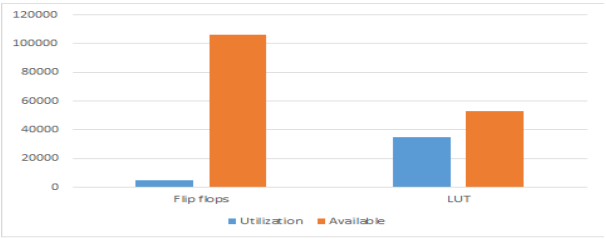


Fig. 8: The logic utilization of the Turbo encoder with serial computation.

The impact of the parallel computation on the logic utilization is shown in Figure 10. Note that the utilized flip flop and LUTs are significantly reduced when parallel computation is employed. The LUT utilization is improved by 73% and the flipflop utilization enhanced by 75%. It is proven that the FPGA technologies strongly supports parallel computation for the Turbo encoder. By reducing the size of the Turbo encoder module, the IVS can be implemented on a single programmable chip with less hardware constrains.

IV. Conclusion

The Turbo encoder module is designed and implemented to be an embedded module in the IVS modem. FPGA technologies are employed to develop the Turbo encoder module. Xilinx tools and Verilog HDL are employed to design and simulate the module. Both serial and parallel computation techniques are studied for the encoding process. It is shown that the parallel computation can improve the chip size and

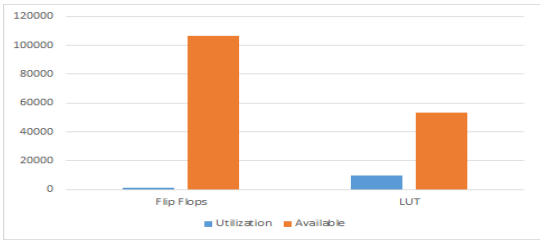


Fig. 9: The logic utilization of the Turbo encoder with parallel computation.

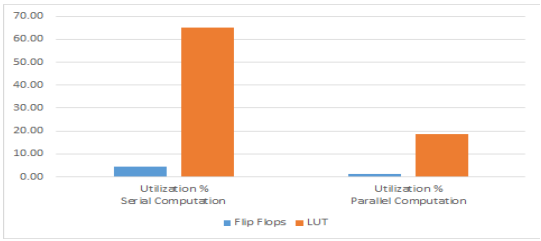


Fig. 10: The logic utilization of the Turbo encoder, serial vs parallel computation.

processing time of the module. Comparing with the serial computation technique, the parallel computation encoding, improves the processing time by 58% and logic utilization by 73%. The processing time enhancement can be seen in both simulation and analyzing the chip processing.

References

[1] "eCall data transfer; in-band modem solution; general description," 3GPP, Tech. Rep. TS26.267.

[2] Eroupean Commission, "eCall: Time saved = lives saved," Press Release, Brussels, August 21, 2015. website: $http://ec.europa.eu/digital-agenda/en/ecall-time-saved-lives-saved$.

[3] A. Saleem et at. "Four-Dimensional Trellis Coded Modulation for Flexible Optical Communications," *IEEE Journal of Lightwave Technology.,* vol. 35, no. 2, pp. 151-158, Nov. 2017.

[4] C. Studer, C. Benkeser, S. Belfanti, and Q. Huang "Design and Implementation of a Parallel Turbo-Decoder ASIC for 3GPP-LTE," *IEEE Journal of Solid-state Circuits.,* vol. 46, no. 1, pp. 8-17, Jan. 2011.

[5] M. Nader and J. Liu, "Design and implementation of CRC module of eCall in-vehicle system on FPGA," *SAE Technical 2015-01-2844,* 2015, doi:10.4271/2015-01-2844.

[6] M. Nader and J. Liu. "Developing modulator and demodulator for the EU eCall in-vehicle system in FPGAs" in *IEEE, 2016 International Conference on Computing, Networking and Communications (ICNC),* Hawaii, USA, Feb. 15-18, 2016, pp. 1-5.

[7] M. Nader and J. Liu. "FPGA Design and Implementation of Demodulator/Decoder Module for the EU eCall In-Vehicle System" in *International Conference on Embedded Systems and Applications (ESA'15),* Las Vegas, USA, July 27-30, 2015, pp. 3-9.

[8] "Technical Specification Group Radio Access Network; Multiplexing and channel coding (FDD)," 3GPP, Tech. Rep. TS22.212.

[9] D. Viktor, K. Michal, and D. Milan"Impact of trellis termination on performance of turbo codes," in *ELEKTRO,* 2016, pp.48-51.

[10] B. Muralikrishna, G.L. Madhumati, H. Khan, K.G. Deepika, "Reconfigurable system-on-chip design using FPGA," in *2nd International Conference on Devices, Circuits and Systems (ICDCS),* 2014, pp.1-6.

[11] J. J. Rodriguez-Andina, M. J. Moure, and M. D. Valdes, "Features, design tools, and application domains of FPGAs," *IEEE Trans. Ind. Electron.,* vol. 54, no. 4, pp. 18101823, Aug. 2007.

Review on Energy Harvesting For Wireless Sensor Networks

P.Venkatapathi [1] Assistant Professor
Bala tejavath[2] Assistant Professor
Godugu Kiran Kumar[3] Assistant Professor

[1,2]Malla Reddy College of Enginnering
[3]Marri Laxman Reddy Institute of Technology and Management

***ABSTRACT* - Sustainable wireless sensor networks (WSNs) are being widely used nowadays due to two key driving technologies behind them i.e. energy harvesting and energy management.**
Energy harvesting from environmental energy sources such as solar, wind, thermal, mechanical and so forth are introduced from the perspective of energy supply to the WSN, while energy management of WSN such as the design of MAC protocol, design of routing protocol, and dynamic power management technology are presented from the perspective of energy conservation within the WSN itself. To better understand them in details for optimizing the sustainable WSN performance, in this paper, a review of these two enabling technologies are performed. More depth research into their combined efforts for sustainable WSN is presented and then illustrated with a case study. One of the most commonly referred energy harvesting source, i.e. solar energy, and its energy management which includes a new energy forecast model of wireless sensor nodes and a new model of energy distribution in WSNs using data collection protocol is investigated and demonstrated.

I. INTRODUCTION

In recent years, wireless sensor networks are widely used in many areas such as disaster management, infrastructure monitoring, security and surveillance, etc [1]. For these applications, the research works are mostly paying attention to the realization of functions in the designing of wireless sensor networks rather than the sustainability issue of the network. The wireless sensor node always uses power-limited battery as its energy supply. However, there are a number of nodes in the wireless sensor networks and they are always distributed in extensively wide and complex environment, it becomes very difficult to change the battery of wireless sensor nodes on deployment [2]. In order to make wireless sensor networks more practical, researchers began to study the sustainability of the wireless sensor networks, namely, try to extend the life cycle of wireless sensor networks effectively [3]. Energy harvesting and energy management are two key technologies that enable a self-sustainable wireless sensor network.

There are many forms of renewable energy readily available in the environment at which the wireless sensor networks are deployed, such as solar energy, mechanical energy, thermal energy, sound energy, wind power and so on. In this paper, we conduct a review of wide varieties of energy harvesting technologies for wireless sensor networks. We mainly focus on how to transform various forms of energy existing in the environment into electrical energy that can be used to sustain the operations of the wireless sensors. Energy management technology is mainly to solve the problem of energy conservation in wireless sensor networks (WSNs) [4] [5]. Energy management usually includes optimization of medium access and routing protocols, dynamic power management etc. However, if solely relying on reducing energy consumption without energy supplement, it is very difficult to maintain long-term operation of a wireless sensor network.

The objective of this paper is to explore on two key enabling technologies of a self-sustainable and self-autonomous wireless sensor network. Firstly, energy harvesting technologies for wireless sensor network, including solar, wind, sound, vibration, thermal, and electromagnetic are introduced. Secondly, energy management technology used in wireless sensor networks are summarized, which include the design of various MAC protocols, routing protocols, cross layer protocols and dynamic power management technology.

Once the individual energy enhancement technology has been explored and researched, the correlation between both of these key technologies is addressed. To be able to fully optimize the WSN to be self-sustainable, rather than just energy harvesting or improved WSN energy management, it is important to further the research discussion into the combination of both energy harvesting and energy management technologies. A case study on the sustainable wireless sensor network that harvests energy from solar power, the energy model of such wireless sensor networks in green building, and the design of its data collection protocol.

II. ENERGY HARVESTING TECHNOLOGY

As we all know, there are many potential uses of stray energies in our living space, such as solar, wind, heat, mechanical vibration, acoustic, electromagnetic energy. These energy sources are free and pollution free. Much research work on large-scale application of environment energy including solar, wind, geothermal, etc. have already been done and the related technologies are very mature [6]. However, when the problem changes into how to harvest and storage these natural energies in small-scale form to power miniaturized wireless sensor nodes, previous large-scale energy harvesting technologies are no longer applicable. Hence, many research works have been discussed in the literature on this energy harvesting technology for self-sustainable wireless sensor network. Some of the key progresses are described as follows.

A. Solar Energy

For earth, solar energy or light energy is a kind of inexhaustible and clean energy. The basic principle of optical collection is to absorb a large number of photons by the use of photovoltaic materials. If there is enough number of photons to activate the electronic optical pool, electricity can be obtained through appropriate structural design. Because power that can be harvested is greatly depending on the light intensity, optical components are usually placed in an environment with good lighting condition in order to obtain more power. Optical components can be connected in serials to generate the required voltage. As manufacturing cost of optoelectronic components is declining, the selection of solar energy as energy source for wireless sensor networks has become a reasonable technical solution.

The only disadvantage of solar energy is that it is only available during day time (for outdoor environment) or office hour (for indoor environment). A battery is needed to ensure the sensors to be operated all around the clock and the efficiency can be low on cloudy days when sun exposure is very low. A number of recent solar energy harvesting prototypes [7]-[9] for sustainable wireless sensor network are presented in Figure 1.

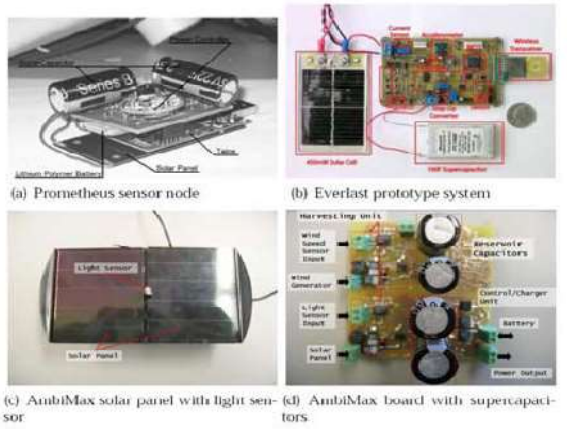


Figure 1: Examples of solar energy harvesting system [7] - [9]

B. Wind Energy

Like any of the commonly available renewable energy sources, wind energy harvesting has been widely researched for high power applications where large wind turbinegenerators (WTGs) are used for supplying power to remote loads and grid-connected applications. Although very few research works are reported in the literature on small-scale wind energy harvesting, some efforts to generate power at a very small-scale have been made recently [10]-[12], and some are presented in Figure 2.

The main disadvantage regarding wind power is unreliability factor, as the strength of the wind is not constant and unpredictable, hence it does

not produce the same amount of electricity all the time. In addition, since it involves moving mechanical part, it can be noisy.

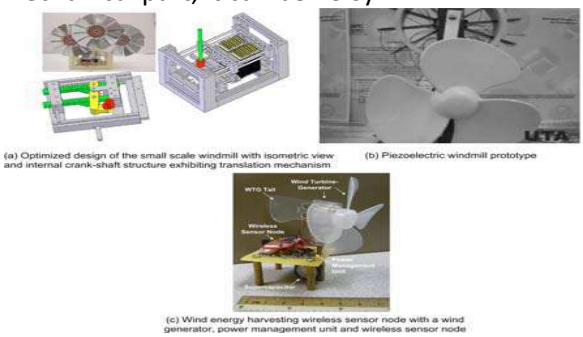


Figure 2: Examples of wind energy harvesting system [10]-[12]

C. Thermal Energy

Research on thermoelectric technology began in 1940's, reached its peak in 1960's. And this technology was successfully used on the spacecraft. Temperature difference generator is featured with characteristics such as small, light weight, no vibration, no noise, less maintenance and can work for long hours under harsh environment. It is suitable to act as low power less than 5W and usually mounted in a variety of unmanned surveillance sensors, tiny short-range communication devices, and medical instrumentation. At present, the relevant products have been widely used. German scientists have invented a new type of battery using the temperature of human body to produce electricity, which can provide long-term "power" for portable miniature electronic devices and eliminates the trouble of charging or replacing batteries. For example, temperature difference which equal to 5 degree C between human skin and clothes can be took advantage of and provide sufficient energy for a common watch. Some examples of the thermal energy harvesting systems are presented in Figure 3.

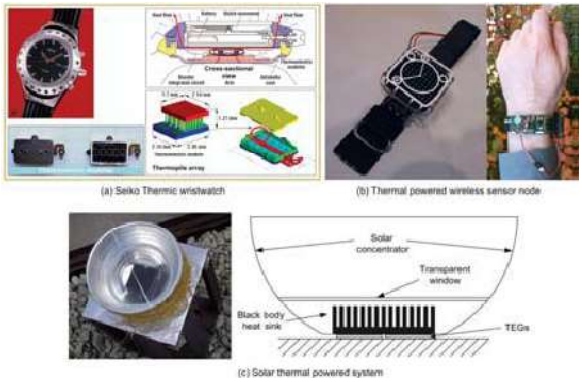


Figure 3: Examples of thermal energy harvesting system [13] -[15]

V. CONCLUSION

In this paper, we provide a comprehensive review on some common energy harvesting technologies of wireless sensor networks, and the introduction of energy management technology. We demonstrate an example of sustainable wireless sensor networks based on solar energy which is for green building. The challenge to harvest environment energy is discussed.

REFERENCES

[1] Y.K. Tan and S.K. Panda (2010). Review of Energy Harvesting Technologies for Sustainable WSN, Sustainable Wireless Sensor Network, Winston Seah (Editor) and Yen Kheng Tan (Editor-in-Chief), ISBN 978-953-307-297-5, INTECH. Available from: http://www.intechopen.com/articles/show/title/review-of-energyharvesting- technologies-for-sustainable-wsn

[2] C. Alippi, G. Anastasi, M. Di Francesco, M. Roveri, “Energy management in wireless sensor networks with energy-hungry sensors”, *IEEE Instrumentation & Measurement Magazine*, vol.12, issue.2, pp.16- 23, 2009.

[3] V. Raghunathan, S. Ganeriwal, M. Srivastava, “Emerging techniques for long lived wireless sensor networks”, *IEEE Communications Magazine*, vol.44, issue.4, pp.108-114, 2006.

[4] C.Y. Chong, S.P. Kumar, “Sensor networks: Evolution, opportunities, and challenges”, *Proceedings of the IEEE Sensor Networks and Applications*, vol.91, no.8, pp.1247-1256, 2003.

[5] Ian F. Akyildiz, Weilian Su, Yogesh Sankarasubramaniam, and Erdal Cayirci, "A Survey on Sensor Networks", *IEEE Communications Magazine*, 2002.

[6] Y.K. Tan and S.K. Panda, "Optimized Wind Energy Harvesting System Using Resistance Emulator and Active Rectifier for Wireless Sensor Nodes", *IEEE Transactions on Power Electronics*, vol.26, issue.1, pp.38- 50, 2011.

[7] X.F. Jiang, J. Polastre and D.E. Culler, "Perpetual environmentally powered sensor networks", *Fourth International Symposium on Information Processing in Sensor Networks (IPSN)*, pp.463-468, 2005.

[8] F.I. Simjee, P.H. Chou, "Efficient charging of supercapacitors for extended lifetime of wireless sensor nodes", *IEEE Transactions on Power Electronics*, vol.23, issue.3, pp.1526-1536, 2008.

[9] C. Park, P.H. Chou, "AmbiMax: Autonomous Energy Harvesting Platform for Multi-Supply Wireless Sensor Nodes", *3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks (SECON)*, vol.1, pp.168-177, 2006.

[10] S.Priya, C.T. Chen, Darren Fye, Jeff Zahnd, "Piezoelectric Windmill: A Novel Solution to Remote Sensing", *Japanese Journal of Applied Physics*, vol.44, no.3, pp.L104L107, 2005.

[11] R. Myers, M. Vickers, Kim Hyeoungwoo, S. Priya, "Small scale windmil", *Applied Physics Letters*, vol.90, no.5, p.54106-1-3, 2007.

[12] Y.K. Tan and S.K. Panda, "Self-Autonomous Wireless Sensor Nodes with Wind Energy Harvesting for Remote Sensing of Wind-Driven Wildfire Spread", *IEEE Transactions on Instrumentation and Measurement*, 2011.

[13] Nicholas S. Hudak and Glenn G. Amatucci, "Small-scale energy harvesting through thermoelectric, vibration, and radio frequency power conversion", *Journal of Applied Physics*, vol.103, no.10, pp.101301(1- 24), 2008.

[14] V. Leonov, T. Torfs, P. Fiorini, C. Van Hoof, "Thermoelectric Converters of Human Warmth for Self-Powered Wireless Sensor Nodes", *IEEE Sensors Journal*, vol.7, issue.5, pp.650-657, 2007.

[15] H.A. Sodano, G.E. Simmers, R. Dereux, D.J. Inman, "Recharging batteries using energy harvested from thermal gradients", Journal of Intelligent Material Systems and Structures, vol.18, no.1, pp.3-10, 2007.

A Simple Yet Efficient Accuracy Configurable Adder Design

Vavaldas Harshini,M.tech scholar,Dept of ECE,MRCE, Hyderabad,harshiniraman143@gmail.com

***Abstract*—Approximate computing is a promising approach for low power IC design and has recently received considerable research attention. To accommodate dynamic levels of approximation, a few accuracy configurable adder designs have been developed in the past. However, these designs tend to incur large area overheads as they rely on either redundant computing or complicated carry prediction. Some of these designs include error detection and correction circuitry, which further increases area. In this work, we investigate a simple accuracy configurable adder design that contains no redundancy or error detection/correction circuitry and uses very simple carry prediction. Simulation results show that our design dominates the latest previous work on accuracy-delay-power tradeoff while using 39% lower area. In the best case the iso-delay power of our design is only 16% of accurate adder regardless of degradation in accuracy. One variant of this design provides finer-grained and largertunability than the previous works. Moreover, we propose a delay-adaptive self-configuration technique to further improve accuracy-delay-power tradeoff. The advantages of our method are confirmed by theapplicationsinmultiplicationandDCTcomputing.**

***Index Terms*—Approximate computing, accuracy configurable adder, delay-adaptive reconfiguration, low power design.**

I. Introduction

POWER constraints are a well-known challenge in advanced VLSI technologies. Low power techniques for the conventional exact computing paradigm have been already extensively studied. A comparatively new direction is approximate computing, where errors are intentionally allowed in exchange for power reduction. In many applications, such as audio, video, haptic processing and machine learning, occasional small errors are indeed acceptable. Such error-tolerant applications are found in abundance in emerging applications and technologies.

A great deal of approximate computing research has been concentrated on arithmetic circuits, which are essential building blocks for most of computing hardware. In particular, several approximate adder designs have been developed [1]–[14]. One such design [2] achieves 60% power reduction for DCT (Discrete Cosine Transform) computation without making any discernible difference to the images being processed. In realistic practice, accuracy requirements may vary for different applications. In mobile computing devices, different powermodesmayentaildifferentaccuracyconstraintsevenfor the same application. Specifically, arithmetic accuracy can be adjusted at runtime using methods such as dynamic voltage and frequency scaling (DVFS) to obtain the best accuracy-power tradeoff. The benefit of runtime accuracy adjustment is demonstrated in [3], but their approximation is realized by voltage over-scaling, where errors mostly occur at the timing-critical path associated with the most significant bits, i.e., errors are oftenlarge.

To reduce the overall error, a few approximate designs have been developed by intentionally allowing errors in lower bits with shorter carry chain in addition operation. In [4], a design that considers only the previous *k* inputs instead of all input bits can approximate the result with the benefit in half of the logarithmic delay. Reliable variable latency carry select adder (VLCSA) shows a speculation technique which introduces carry chain truncation and carry select addition as a basis [7]. A series of Error Tolerant Adders (ETAI, ETAII, ETAIIM), which truncate the carry propagation chain by dividing the adder into several segments, have been proposed [8]–[10]. Correlation-aware speculative adder (CASA) in [11] relies on the correlation between MSBs of input data and carry-in values. Another approximate adder that exploits the generate signals for carry speculation is presented [12]. These designs focus on static approximation which pursues almost correct results at the required accuracy. However, in some applications such as image processing or audio/video compression, the required accuracy might vary during runtime. To meet the need of runtime accuracy adjustment, a series of designs are developed to implement accuracy configurable approximation whichcouldbereconfiguredonlinetosavemorepower.

A few accuracy configurable adder designs that use approximation schemes other than voltage over-scaling have been proposed. An early work [15], called ACA, starts with an approximate adder and augments it with an error detection and correction circuit, which can be configured to deliver varying approximation levels or accurate computing. Its baseline approximate adder contains significant redundancy and the error detection/correction circuit further increases area overhead. The ACA design [15] is generalized to a flexible framework GeAr in [16]. In both ACA and GeAr, the error correction must start from the least significant bits and hence accuracy improves slowly in the progression ofconfigurations. The work of Accurus [17] modifies ACA/GeAr to overcome this drawback and achieves graceful degradation. However, in ACA, GeAr as well as Accurus, the error correction circuit is pipelined, implying that the computation in accurate mode takesmultipleclockcyclesandcausesdatastalls.

This work was partially supported by NSF (CCF-1255193, CCF-1525749 and CCF-1525925).

W. Xu and J. Hu are with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, 77843 USA (e-mail: wbxu@tamu.edu; jianghu@tamu.edu).

S. S. Sapatnekar is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, 55455 USA (email: sachin@umn.edu).

An alternative direction of accuracy configurable adder design is represented by GDA [18] and RAP-CLA [19]. These methods start with an accurate adder and use carry prediction for optional approximation. As such, they no longer need error detection/correction and do not incur any data stall. In addition, they intrinsically support graceful degradation. The GDA design [18] is composed by accurate CRA (Carry Ripple Adder) and extra configurable carry prediction circuitry, similar as the carry look-ahead part of CLA (Carry Look-ahead Adder). Thus, its area is generally quite large. RAP-CLA [19] is based on accurate CLA design and reuses a portion of the carry look-ahead circuit as carry prediction. This leads to an overall area that is less than GDA but greater than CLA. In [19], the carry-prediction-based approach is shown to be superior to error-correction-based design[16].

In this paper, we propose a new carry-prediction-based accuracy configurable adder design: **SARA** (Simple Accuracy Reconfigurable Adder). It is a simple design with significantly less area than CLA, which, to the best of our knowledge, has not been achieved in the past in accuracy configurable adders. SARA inherits the advantages of all previouscarry-prediction-basedapproaches:noerrorcorrectionoverhead,nodatastall andallowinggracefuldegradation.ComparedtoGDA[18], SARA incurs 50% less PDP (Power Delay Product) and can reach the same PSNR (Peak Signal-to-Noise Ratio). Moreover, SARA demonstrates remarkably better accuracy-power-delay tradeoff than the latest, and arguably the best, previous work RAP-CLA [19]. A delay-adaptive reconfiguration technique is developed to further improve the accuracy-power-delay tradeoff. The proposed designs are also validated by multiplication and DCT computation in image processing.

II. Prior Works and Rationale of Our Design

Wereviewafewrepresentativeworksonaccuracyconfigurable adder design and show the relation with our method. These designs can be generally categorized into two groups: error-correction-based configurations [15]–[17] and carry-prediction-based configurations [18], [19].

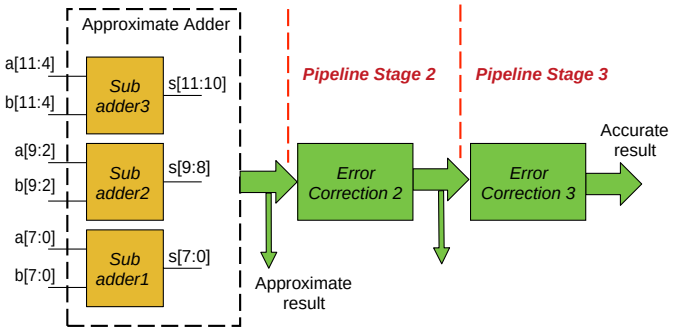


Fig. 1. Error-correction-based configurable adder.

Themainideaofanerror-correction-basedapproach[15]–[17] is shown in Figure 1. The scheme starts with an approximate adder (the dashed box), where the carry chain is shortened by using separated sub-adders with truncated carry-in.Inordertoreducethetruncationerror,thebit-width in some sub-adders contains redundancy. For example, *sub-adder2* calculates the sum for only bit 8 and 9, but it is an 8-bit adder using bit $[9:2]$ of the addends, 6 bits of which are redundant. Even with the redundancy, there is still residual error which is detected and corrected by additional circuits. In Figure 1, the errors of *sub-adder2* must be corrected by *error-correction2* before the errors of *sub-adder3* are rectified by *error-correction3*. As such, the configuration progression always starts with small accuracy improvements. The redundancy and error detection/correction incur large area overhead. Since the error correction circuits are usually pipelined, an accuratecomputationmaytakemultipleclockcyclesandcould stallentiredatapath,dependingontheaddendvalues.

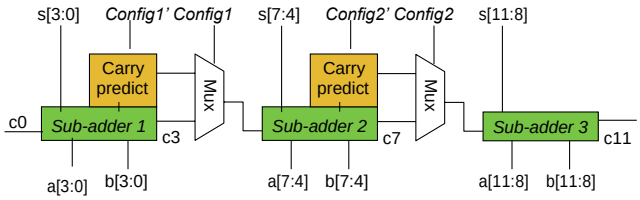


Fig. 2. Carry-prediction-based configurable adder.

Theframeworkofcarry-prediction-basedmethods[18],[19] is shown in Figure 2. These schemes start with an accurate adder design, which is formed by chaining a set of sub- adders. Each sub-adder comes with a fast but approximated carry prediction circuit. By selecting between the carry-out from sub-adder or carry prediction, the overall accuracy can be configured to different levels. Such an approach does not need error detection/correction circuitry. Moreover, the configuration of higher bits is independent of lower bits. This leads to fast convergence or graceful degradation in the progression of configurations. In GDA [18], the sub-adders are CRA designs while the carry-prediction circuit is similar tothe carry look-ahead part of CLA. Further, its carry prediction can be configured to different accuracy levels. However, the complicated carry prediction induces large area overhead. The RAP-CLA scheme [19] uses CLA for its baseline where the carry-ahead of each bit is computed directly from theaddends of all of its lower bits. Its carry prediction reuses a part of the look-ahead circuit rather than building extra dedicated prediction circuitry, and hence is more area-efficient than GDA.ButitsbaselineismuchmoreexpensivethanGDA.

TABLE I
Comparison of characteristics for different techniques.

Method	Baseline sub-adder	Error correction	Graceful degradation	Carry prediction
ACA [15]	Redundant CRA	Yes	No	No
GeAr [16]	Redundant CRA	Yes	No	No
Accurus [17]	Redundant CRA	Yes	Yes	No
GDA [18]	CRA	No	Yes	Stand-alone
RAP-CLA [19]	CLA	No	Yes	Reuse
SARA (ours)	CRA	No	Yes	Reuse

Our design is a carry-prediction-based approach. Its sub-adders are CRA instead of expensive CLA as in RAP-CLA. Itscarrypredictionalsoreusespartofthesub-addersrather

than having dedicated prediction circuitry. As such, it avoids the disadvantages of both GDA and RAP-CLA. A comparison among the characteristics of these different techniques is provided in Table I.

III. Simple Accuracy Reconfigurable Adder

A. Preliminaries

An N-bit adder operates on two addends $A = (a_N, a_{N-1}, ..., a_i, ..., a_1)$ and $B = (b_N, b_{N-1}, ..., b_i, ..., b_1)$. For bit i, its carry-in is c_{i-1} and its carry-out is c_i. Defining the carry generate bit $g_i = a_i b_i$, propagate bit $p_i = a_i \oplus b_i$ and kill bit $k_i = \bar{a}_i \bar{b}_i$, the conventional full adder computes the sum s_i and carry c_i according to

$$s_i = p_i \oplus c_{i-1}, \tag{1}$$

$$c_i = g_i + p_i \cdot c_{i-1}. \tag{2}$$

A gate level schematic of conventional full adder is provided in Figure 3(a). A CRA is used to chain N bits of conventional full adders together.

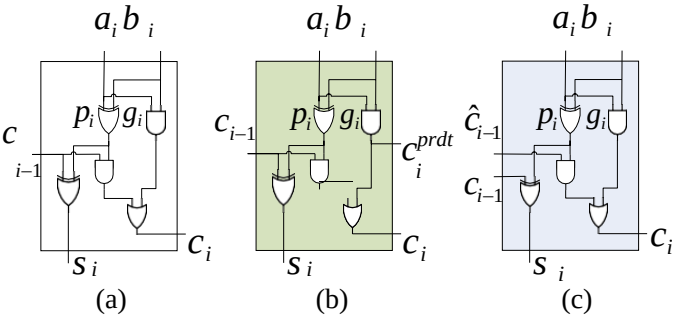


Fig.3. (a) Conventional full adder; (b) Our carry-out selectable full adder; (c) Our carry-in configurable full adder.

By applying Equation (2) recursively, one can get

$$c_i = g_i + p_i g_{i-1} + ... + g_1 \prod_{k=2}^{i} p_k + c_0 \prod_{k=1}^{i} p_k. \tag{3}$$

This equation implies that c_i can be computed directly from g and p of all bits, without waiting for the c of its lower bits to be computed. This observation is the basis for CLA adder.

B. SARA: Simple Accuracy Reconfigurable Adder Design

In SARA, an N-bit adder is composed by K segments of L-bit sub-adders, where $K = \lfloor N/L \rfloor$ (see Figure 2). Each sub-adder is almost the same as CRA except that the MSB (Most Significant Bit) of a sub-adder, which is bit i, provides a carry prediction as

$$c_i^{prdt} = g_i \tag{4}$$

For the LSB (Least Significant Bit) of the higher-bit sub-adder, which is bit $i + 1$, its carry-out c_{i+1} can be computed using one of two options: either by the conventional $c_{i+1} = g_{i+1} + p_{i+1} \cdot c_i$, or by using the carry prediction as

$$c_{i+1} = g_{i+1} + p_{i+1} \cdot c_i^{prdt} = g_{i+1} + p_{i+1} \cdot g_i \tag{5}$$

The selection between the two options is realized using MUXes as in Figure 4 and the MUX selection result is denoted as $\hat{c}_i$. Comparing Equation (5) with (3), we can see that the carry prediction is a truncation-based approximation to carry computation[1]. Therefore, $\hat{c}_i$ can be configured to either accurate mode or approximation mode, i.e.,

$$\hat{c}_i \leftarrow \begin{cases} c_i^{prdt}, & \text{if approximation mode} \\ c_i, & \text{if accurate mode.} \end{cases} \tag{6}$$

It should be noted that the carry prediction c_i^{prdt} reuses g_i in an existing full adder instead of introducing an additional dedicated circuit as in [18] or Figure 2. This prediction scheme makes a very simple modification to the conventional full adder, as shown in Figure 3(b).

One can connect $\hat{c}_i$ to its higher bit $i+1$ to compute both carry c_{i+1} and sum s_{i+1}, as in GDA [18] and RAP-CLA [19]. We suggest an improvement over this approach by another simple change as in Figure 3(c), where s_{i+1} is based on c_i instead of $\hat{c}_i$. Such approach can help reduce the error rate in outputs when an incorrect carry is propagated. Because the sum keeps accurate and the carry will not be propagated when addends are exactly the same. Moreover, out of all four configurations of sum/carry calculation by approximate/accurate carry-in, the most meaningful way is to have sum bit calculated by accurate carry and make carry bit configurable. So sum s_{i+1} is calculated directly by accurate carry c_i without the option of c_i^{prdt}. Applying this in SARA as in Figure 4 in the approximation mode, computing s_{j+1} from c_j can still limit the critical path to be between c_{i-1} and s_{j+1}, but has higher accuracy than computing s_{j+1} from $\hat{c}_j$. Compared to sum computation in GDA and RAP-CLA, this technique improves accuracy with almost no additional overhead. Compared to CRA, the overhead of SARA is merely the MUXes, which is almost the minimum possible for configurable adders.

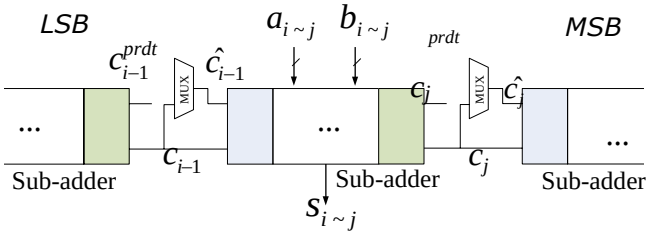


Fig. 4. Design of SARA.

Although s_{j+1} is calculated by accurate carry c_j, its delay can still be reduced by approximate carry in lower sub-adder. In a multi-bit adder, the delay of sum bit depends on the carry chain propagated from its lower bits. In our SARA structure, even when accurate carry c_j is propagated at bit j, the carry chain might be truncated by approximate carry in other lower bits. In Figure 4, when c_{i-1}^{prdt} is propagated, the delay of s_{j+1} is reduced as its path is shorten to be between bit $i-1$ and

[1] A similar approximation is used in static approximate adder design [12].

$j+1$. We can take the 12-bit adder in Figure 5 as an example. For 12-bit SARA working in approximate mode, the sum $s_9$uses the accurate carry $c_8$from a lower sub-adder (bit 5 to 8). But $c_8$is propagated from approximate carry $c_4^{prdt}$of another sub-adder (bit 1 to 4). As shown in the figure, the delay of $s_9$in SARA is about 6 stages. Compared with the same bit in CRA, the delay of sum bit $s_9$in SARA is reduced by 3 stages. Similar delay reduction can be observed in other sum bits (bit 6 to 12). For sums at bit 1 to 5, their delay is the same as CRA because they are using an accurate carry $c_0$from LSB. As a result, the maximum delay in 12-bit SARA is reduced, since for a multi-bit adder its maximum delay depends on the longest criticalpath.

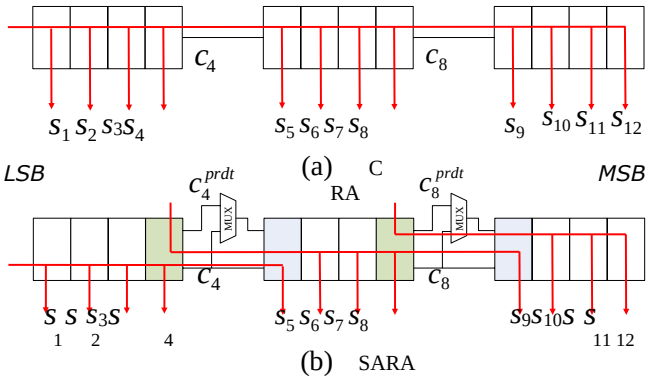


Fig. 5. Implementation of 12-bit adder in (a) CRA and (b) SARA.

C. Usage ofSARA

When$\hat{c}_i$isconfiguredtobe$c_i$forall$K$sub-adders,SARA operates very much like the CRA, where the critical path is along$N$-bitfulladders.Ifall$\hat{c}_i$areselectedtobec^{prdt},the $_i$ critical path is shortened to roughly L-bit full adders. This large delay reduction can be translated to power reduction by supply voltage scaling. Voltage scaling (reducing supply voltage) on digital circuits will lead to increase in delay. So wecanreducethesupplyvoltageonSARAtomakeitscritical delay same as that of CRA under normal voltage. As the supply voltage decreases, the power consumption could be reduced.Therecan be2^{K-1}differentconfigurations.Fortwo configurations with the same critical path length, obviously we only need the one with higher accuracy. Therefore,there are$K$effectiveconfigurations,withcriticalpathlengthof$L$- bit, $2L$-bit, ..., $K\ L\cdot N$ bit full adders. The delay of such configurabledesignvariesaccordingtoconfiguredaccuracy, whichresultsindifferentpowerreductionbyvoltagescaling.

IV. SARA ErrorAnalysis

Inthissection,wegiveatheoreticalanalysisontheexpected errorofourSARAdesignandvalidatetheresultsbynumerical experiments. To make it easier for readers to follow the analysis, we list the parameters used in this section as TableII.

For any bit i in carry-out selectable full adder as in Figure3(b),anerrorinapproximatecarry-outoccurswhen

TABLE II
DEFINITION OF PARAMETERS FOR ERROR ANALYSIS

Parameter	Definition
p_i	propagate bit at bit i
g_i	generate bit at bit i
k_i	kill bit at bit i
c_i	accurate carry-out bit at bit i
c_i^{prdt}	approximate carry-out bit at bit i
$\hat{c}_i$	carry-in bit at bit $i+1$
ER_i^{prdt}	error rate of c_i^{prdt}
$\widehat{ER}_i$	errorrateof$\hat{c}_i$

c_i^{prdt} c_i. There is only one situation where this error may happen: when $c_{i-1}=1$, $p_i=1$, $c_i^{prdt}=0$ and $c_i=1$. Then the error rate, or probability of such error, is given by

$$\begin{aligned} ER_i^{prdt} &= P\,(c_i^{prdt} \neq c_i) = P\,(c_i^{prdt}=0,\ c_i=1) \\ &= P\,(c_{i-1}=1,\ p_i=1) \\ &= P(c_{i-1}=1)P(p_i=1) \end{aligned} \tag{7}$$

where P indicates probability and the last part assumes that $c_{i-1}$and $p_i$are independent of each other. Then, if the approximate/accurate carry-out can be selected by a MUX gate,theerrorrateofMUXoutput$\hat{c}_i$is

$$\widehat{ER}_i = P(\hat{c}_i \neq c_i) = \begin{cases} ER_i^{prdt}, & \text{if } \hat{c}_i \leftarrow c_i^{prdt} \\ 0, & \text{if } \hat{c}_i \leftarrow c_i. \end{cases} \tag{8}$$

Let'sconsideraconfigurationofSARAinFigure4,which has both bit j and bit i 1 in approximate mode. For the sub-adder which calculates addends from bit i to bit j, its LSB(bit$i$)isusingcarry-inconfigurablefulladder,whileits MSB(bit$j$)isincarry-outselectablefulladder.Accordingto Equation(7)and(8),theerrorrateof$\hat{c}_j$isdeterminedbythe probabilities of $c_{j-1}=1$ and $p_j=1$.

$$\widehat{ER}_j = P\,(c_{j-1}=1)P(p_j=1) \tag{9}$$

Accordingtothelogicofaddition,thecarry-outbitiscalculatedbythecarry-inandaddends.Therearetwocaseswhich can result in $c_{j-1}=1$: generate bit g_{j-1}should be 1 in case of carry-in $c_{j-2}=0$; or kill bit k_{j-1}must be 0 when carry-in comes with $c_{j-2}=1$. Then, the probability of $c_{j-1}=1$ can be computed by the probability of $c_{j-2}=1$ as

$$\begin{aligned} P(c_{j-1}=1) &= P(c_{j-2}=0, g_{j-1}=1) + P(c_{j-2}=1, k_{j-1}=0) \\ &= P(c_{j-2}=0)P(g_{j-1}=1) + P(c_{j-2}=1)P(k_{j-1}=0) \\ &= [1-P(c_{j-2}=1)]P(g_{j-1}=1) + P(c_{j-2}=1)P(k_{j-1}=0). \end{aligned} \tag{10}$$

Similarly,theprobabilityof$c_{j-2}=1, c_{j-3}=1, \ldots, c_{i+1}=1$ canbecalculatedusingthesameformula.Fortheprobability of$c_i=1$,it'salittledifferentbecausethecarry-out$c_i$inour carry-inconfigurablefulladderisbasedonpredictedcarry-in $\hat{c}_{i-1}$insteadofc_{i-1}.Consideringthatbit$i$1isconfigured in approximate mode, wehave

$$P(\hat{c}_{i-1}=1) = P(c_{i-1}^{prdt}=1) = P(g_{i-1}=1). \tag{11}$$

Then, the probability of $c_i=1$ can be expressed as

$$P(c_i=1) = [1 - P(g_{i-1}=1)]P(g_i=1) + P(g_{i-1}=1)P(k_i=0). \tag{12}$$

TABLE III
ERROR RATE OF SUB-ADDER WITH DIFFERENT WIDTH

Sub-adder length L	Calculated error rate	Simulated error rate
1	1/8 = 0.125	0.1257
2	3/16 = 0.1875	0.1879
3	7/32 = 0.21875	0.2187
4	15/64 = 0.234375	0.2347
5	31/128 = 0.2421875	0.2424
6	63/256 = 0.24609375	0.2464

By expanding Equation (10) recursively till bit i, the probability of $c_{j-1}=1$ can be calculated by a function of generate bit and kill bit from bit $i-1$ to bit $j-1$.

$$P(c_{j-1}=1)=f\{P(g_{i-1}=1),...,P(g_{j-1}=1), P(k_i=0),...,P(k_{j-1}=0)\}. \quad (13)$$

Assuming that the inputs for adder are uniformly distributed random numbers, we have $P(g=1)=1/4, P(k=0)=3/4$. As the length of sub-adders varies from 1 to 6, the error rates of $\hat{c}_j$ calculated by Equation (9) are listed in the second column of Table III. Corresponding data from numerical simulation in Matlab are also presented in the last column. The error rates calculated by our method match well with experiment results, which demonstrates the correctness of our mathematical analysis. We can also observe that as the length of sub-adder increases the error rate is bounded by 0.25. That is because when the length of sub-adder comes to infinite the probability of $c=1$ will become 0.5 as the normal carry in accurate adder.

Theorem 1. *If $\mathbf{I}$ is the set of bits with MUX at output, the expected error of SARA for unsigned integers is*

$$\sum_{\forall i\in \mathbf{I}} \widehat{ER}_i \cdot P(p_{i+1}=1)\cdot 2^{i+1}.$$

Proof. The overall expected error of SARA can be calculated by summing respective error introduced by every approximate bit from LSBs to MSBs. But the propagation of inaccurate carry bit may cause error in higher bit which also be counted in the calculation of lower bit. So we need to exclude those errors to avoid over-calculation in the total error.

Let's consider the SARA design in Figure 4 which have approximate configuration at both bit $i-1$ and bit j. Assuming that bit $i-1$ is the lowest bit configured in approximate mode, we know that all sum bits $s_k (k \in [1, i-1])$ as well as carry bit c_{i-1} are accurate.

$$c_{i-1} = c_{i-1}^{acc} \quad (14)$$

Then the probability that carry prediction at MUX output $\hat{c}_{i-1}$ mismatches with accurate carry c^{acc} should be the same as the error rate of max ouput $\hat{c}_{i-1}$.

$$P(\hat{c}_{i-1} \; c_{i-1}^{acc}) = P(\hat{c}_{i-1} \neq c_{i-1}) = \widehat{ER}_{i-1} \quad (15)$$

According to the structure of carry-in configurable full adder (Figure 3(c)), sum bit s_i calculated from c_{i-1} is always accurate; however, the carry-out bit c_i becomes conditionally accurate which depends on both carry-in bit and propagate bit. As shown in Equation (16), the scenario of accurate carry-out can be attributed to two conditions: when the carry-in is not accurate, the carry-out bit becomes accurate as the propagate bit is false; otherwise, it must be accurate no matter what kind of addends are given.

$$P(c_i=c_i^{acc}) = P(\hat{c}_{i-1}=c_{i-1}^{acc}) + P(\hat{c}_{i-1} \; c_{i-1}^{acc}) P(p_i=0) \quad (16)$$

Its complementary part, the probability of inaccurate carry c_i, can be expressed as

$$\begin{aligned} P(c_i \; c_i^{acc}) &= P(\hat{c}_{i-1} \; c_{i-1}^{acc}) P(p_i=1) \\ &= P(\hat{c}_{i-1} \; c_{i-1}) P(p_i=1) \\ &= \widehat{ER}_{i-1} \cdot P(p_i=1). \end{aligned} \quad (17)$$

As a result, the approximation at bit $i-1$ would cause an inaccurate carry-in c_i at bit $i+1$, which introduces the magnitude of 2^i to the overall error in final result. Then the expected error introduced by approximation at bit $i-1$ can be estimated by

$$E[e_{i-1}] = P(c_i \neq c_i^{acc}) \cdot 2^i = \widehat{ER}_{i-1} \cdot P(p_i=1) \cdot 2^i. \quad (18)$$

Next, we consider the expected error introduced by approximation at bit j. As bit j is not the lowest bit in approximate mode, there is a chance that the propagation of inaccurate carry from bit $i-1$ induces error at bit j while it has be taken into account in the error calculation of bit $i+1$. Then the problem is whether the carry c_j is accurate when there is a mismatch between $\hat{c}_j$ and c_j. If not, we need to exclude the impact from lower bit when estimating the error at bit j. Let's answer this question in the following cases.

- Case 1: If any propagate bit in sub-adder (bit i to j) equals 0, the error propagation by inaccurate carry will be paused. In another word, the error carried by inaccurate carry bit cannot be propagated to higher bit any more, because the carry-out is independent of carry-in when propagate bit is false. In this case, the carry c_j should be always accurate regardless of the configuration at bit j.
- Case 2: If all propagate bits of sub-adder equal 1, the value of inaccurate carry $\hat{c}_{i-1}$ (0 instead of 1) will be propagated to c_j. In this situation, the actual value of c_j propagated from bit $i+1$ must be 0, while the accurate value should be 1. Assuming that $\hat{c}_j$ mismatches with c_j, we can state that the value of $\hat{c}_j$ must be 1. However, it conflicts with the generation of c_j, because carry c_j is the logical conjunction of $\hat{c}_j$ and $p_j c_{j-1}$. So there should be no mismatch between $\hat{c}_j$ and c_j in this case.

In conclusion, when there is a mismatch between $\hat{c}_j$ and c_j, the value of carry c_j must be accurate. We can further conclude that the contributions of every approximate bit to the total error are independent to each other. Similar to bit $i-1$, the expected error at bit j can be estimated by

$$E[e_j] = \widehat{ER}_j \cdot P(p_{j+1}=1) \cdot 2^{j+1}. \quad (19)$$

Thus, the total error can be obtained by summing up the errors respectively introduced by every approximate bit.

$$E = \sum_{\forall i\in \mathbf{I}} E[e_i] = \sum_{\forall i\in \mathbf{I}} \widehat{ER}_i \cdot P(p_{i+1}=1) \cdot 2^{i+1} \quad (20)$$

□

If input addends are random variables following uniform distribution, the expected error of SARA is given by

$$E = \sum_{\forall i \in I} \widehat{ER}_i \cdot 2^i \quad (21)$$

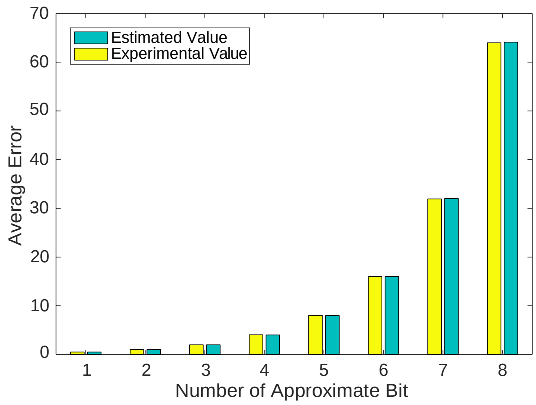


Fig. 6. Average error of 9-bit SARA in different configuration.

We can verify Equation (21) by numerical simulation of a 9-bit SARA design. In our experiment, SARA consists of 9 sub-adders whose width is 1 bit. The results are from 200K run of Monte Carlo simulation with uniform distributed numbers as input. As shown in Figure 6, there are 2 sets of data for comparison, experimental data are obtained directly in experiments and estimated data are calculated by Equation (21). The average errors from experiment are almost the same as the estimated values. According to the analysis above, we can estimate the average error of SARA in any configuration, given the distribution of input numbers.

Since $|I|\,K\,1$, the error of the worst case approximation mode increases with the number of sub-adders, K. In addition, area overhead increases with K. On the other hand, a large K implies smaller L, and thus often facilitates shorter critical path and more power reductions. Therefore, K significantly affects the tradeoff among accuracy, power, delay and area.

V. Delay-Adaptive Reconfiguration of SARA

Almost all previous works on accuracy configurable adder [15]–[19] reasonably assume that accuracy configuration is decided by architecture/system level applications. We propose a self-configuration technique for the scenarios where architecture/system level choice is either unclear or difficult. Simulation results show that SARA with the self-configuration outperforms several previous static approximate adder designs. The main idea of self-configuration is based on the observa-tion that the actual worst case path delay depends on addend values. Specifically, the actual path delay is large only when a carry is propagated through several consecutive bits. Any false propagate bit from the addends results in a shorter carry propagation chain. When the actual carry propagation chain is short, there is no need to use approximation configuration, which is intended to cut carry chain shorter. We propose a Delay Adaptive Reconfiguration (**DAR**) technique: the output of a MUX in SARA is set to approximation mode only when a potentially long carry chain is detected. Compared to the constantly-approximate configuration, some errors for actual short carry chains are avoided, the actual long carry chain is cut shorter, and delay/power reduction can be still obtained.

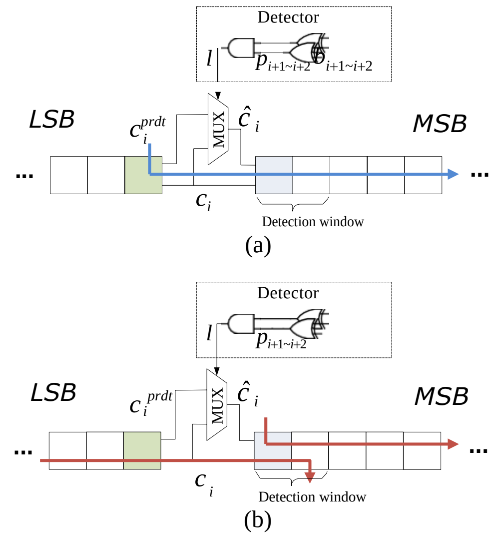


Fig. 7. Design and operation of delay-adaptive reconfiguration for SARA.

The long carry chain detection and SARA-DAR design are shown in Figure 7(a). When MUX is switched to accurate mode by any false propagate bit in detection window, the actual carry chain is retained by the position of false propagate bit. To obtain a shorter carry chain in accurate mode, the detection window for MUX at bit i in MSB should start from bit i+1. In the example of Figure 7, we use a detection window of 2 bits (p_{i+1} and p_{i+2}) to tell if there is a carry propagation across two sub-adders, and configure the MUX according to

$$\hat{c}_i \leftarrow \begin{cases} c_i^{prdt}, & \text{if } p_{i+1} \cdot p_{i+2} \text{ is true} \\ c_i, & \text{otherwise.} \end{cases} \quad (22)$$

In approximation mode, the effective carry chain is represented by the blue line in Figure 7(a) and its length is no greater than $L+1$ bits. When the MUX is set to accurate mode, the carry chain is indicated by the red lines in Figure 7(b) and their lengths can be restrained to within $L+2$ bits. Since the propagate bits only depend on local primary inputs, we can reuse propagate bits in higher bits to save cost. Note that in this case the detection overhead here is almost the minimum possible, i.e., only one NAND gate for configuring each MUX. In Figure 7, we use 2-bit detection window, which can be generalized to W-bit. Then, the error rate for MUX at bit i becomes

$$\widehat{ER}_i^{dar} = ER_i^{prdt} \cdot \prod_{j=1}^{W} P(p_{i+j} = 1) \quad (23)$$

The detection window size W decides the tradeoff between accuracy and the effective carry chain length in accurate mode, which is $L+W$. When W increases, the error rate decreases while the critical path length in accurate mode increases.

VI. Experimental Results

A. Experiment Setup and Evaluation

Our SARA, SARA-DAR and several previous designs are synthesized to 32-bit adders by Synopsys Design Compiler using the Nangate 45nm Open Cell Library. The synthesized circuits are placed and routed by Cadence Encounter. The default supply voltage level is 1.25V. To make fair comparisons across architectures, we describe all designs bystructural modeling in Verilog to reduce the impact of synthesis and optimization. For comparison, we synthesize the accurate adder in behavioral modeling which is described by expressional operator in Verilog. The netlist of such accurate adder should be automatically optimized by synthesizer in Design Complier, which is different from any man-craft gate-level design. In addition, we set the same supply voltage and no delayconstraintonalldesignsforthesamereason.

The evaluation of accuracy configurable adder designs can be subtle and therefore is worth some discussion.

1) **Area**: In the literature, the area sometimes refers to the part of the circuit working in a certain mode, e.g., the circuit for the accurate part is not included in area estimation when evaluating approximation mode. We reporttheroutedlayoutareaofeachentiredesign.
2) **Delay**: Some configurable adders, such as ACA [15] and GeAr [16], implement error correction with pipelining, which sometimes takes multiple clock cycles to determine the complete result. The delay or performance evaluation of such designs is much more complicated than unpipelined designs. Our work is focused on unpipelined implementation, although it can be pipelined. Thus, the reported delay is the maximum combinational logic path delay obtained from Synopsys PrimeTime with consideration of wiredelay.
3) **Power**: The power dissipation is estimated by Synopsys PrimeTimeconsideringbothstaticanddynamicpower.
4) **Accuracy**: We use PSNR (Peak Signal-to-Noise Ratio), where errors are treated as noise, as a composite accuracy metric for considering both error magnitude and error rate. In addition, the worst case error, which is equivalent to the maximum error magnitude [13], and error rate are also reported. Each error result is from 100K-run Matlab-based Monte Carlo simulation assuming uniform distribution ofaddends.
5) **Tunability**: This means the range and granularity of runtime accuracy configurations. Sometimes, this can be confused with design-timeflexibility.
6) **Tradeoff**: The tradeoff among the above factors is complex and is difficult to capture in a simple picture. To this end, we use composite metrics including power-delay product (PDP), energy-delay product (EDP) and iso-delaypower.

B. Results of Tradeoff for DifferentConfigurations

In this part, we mainly compare the following accuracy configurable adder designs:

- GDA [18]: We use the same design as in [18], whereeachsub-adderhas4bits.Thisdesigncanbeconfigured by choosing accurate or predicted carry-out for each sub-adder. The carry prediction at each segment can also be configured to different accuracy levels by using different number of lower-bit addends.
- RAP-CLA [19]: We implement four different designs with carry prediction bit-width from 1 bit to 4 bits, which is reflected in the name. For example, RAP-CLA2 means each of its carry prediction is from its 2 lower bits. As in [19], each design can be configured to either only one approximation mode or accuratemode.
- SARA: This is our proposed design and we evaluate sub-adder bit-width of 1 bit, 4 bits and 8 bits, referred to as SARA1, SARA4 and SARA8,respectively.

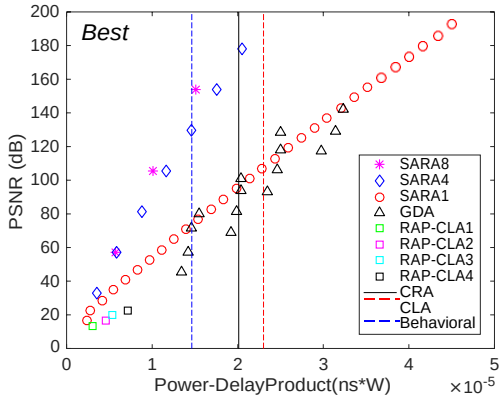


Fig. 8. SARA: PSNR versus power-delay product.

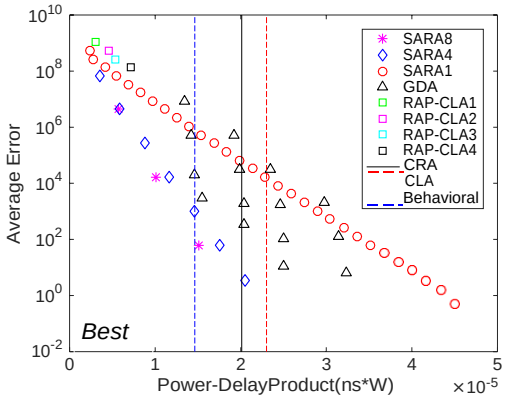


Fig. 9. SARA: Average error versus power-delay product.

The main result is shown in Figure 8, where each point is from one configuration of one design. The computation accuracy is evaluated by PSNR while the conventional design objectives are characterized by PDP. A design and configuration is ideal if it has large PSNR but low PDP, i.e., northwest in the figure. PDPs of two classic accurate designs, CRA and CLA, are indicated by the two vertical lines as their PSNR is near infinity. The result of SARA working in completely accurate mode is unable to be presented in the figure, because its infinite PSNR cannot be displayed as a single dot in theplot.Evidently,thebestsolutionsarefromSARA4and

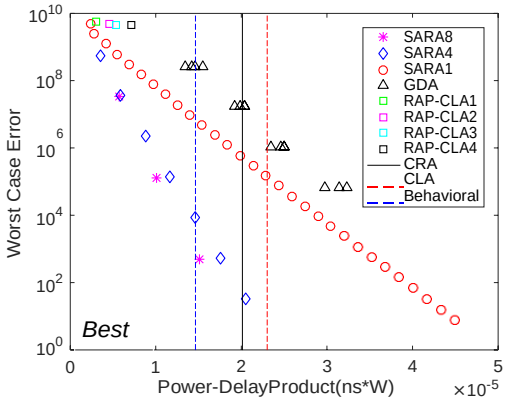


Fig.10. SARA:Theworstcaseerrorversuspower-delayproduct.

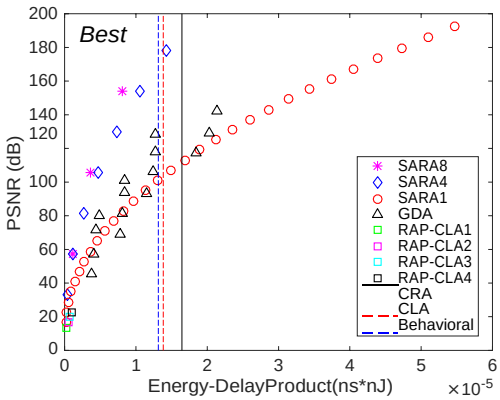


Fig. 11. SARA: PSNR versus energy-delayproduct.

SARA8. At 100*dB* PSNR, the PDP of SARA4 and SARA8 is about a half of GDA or CRA. The solutions from RAP-CLA, the latest previous work, are also largely dominated by SARA in PSNR-PDP tradeoff. An interesting case is SARA1. Its tradeoff is similar as GDA and not as good as SARA4 or SARA8. However, its runtime tunability is superior to all the other designs. It has the largest tuning range, the finest tuning granularity and very smooth tradeoff.

Figure 9 and 10 show the tradeoff between error magnitude and power-delay product. Ideally a better design or configuration has smaller average error or worst case error with lower PDP, which can be marked in the lower left corner of thefigure.InFigure9,SARA4andSARA8dominateother designs in average error-PDP tradeoff. For each configuration, SARA4 and SARA8 have almost the lowest average error at a certain PDP level. Although SARA1 cannot achieve superior average error and PDP tradeoff to GDA, it shows fine-grant tunability in a large range same as PSNR-PDP tradeoff. Figures 10 depicts the worst case error versus PDP and confirms the trend observed in the PSNR-PDP tradeoff. All SARA designs even for SARA1 have lower worst case error than previous work at the same PDP level. In addition, the result of SARA working in accurate mode cannot be found in the plot. That's because the y-axis is in Logarithmic scale and zero error will be converted into infinite which cannot be displayed as a singledot.

EDP is another metric to efficiently evaluate tradeoffs betweencircuitlevelpowersavingtechniquesfordigitaldesigns. Figure 11 to 13 illustrate accuracy versus EDP, which have similar trend in accuracy-PDP tradeoff. Most configurations of SARA4 and SARA8 have lower EDP than accurate adder CRA and CLA. At a certain EDP level, SARA4 and SARA8 still dominate GDA and RAP-CLA with larger PSNR, smaller average error or worst case error. SARA1 in different configurations cover the range from lowest to highest EDP, which provides finest tunability in accuracy-energy tradeoff among differentarchitectures.

C. Results of Tradeoff for Delay-AdaptiveReconfiguration

This part is to evaluate the SARA-DAR design, where the configuration decision has already been made. Hence, it makes

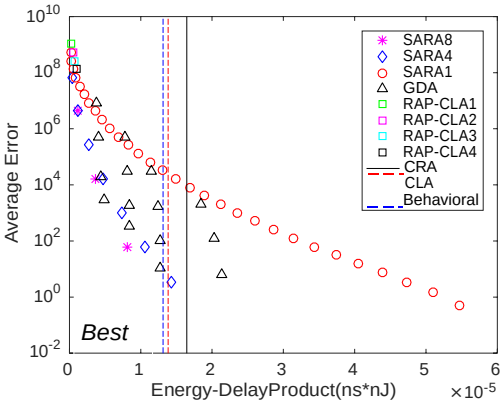


Fig. 12. SARA: Average error versus energy-delay product.

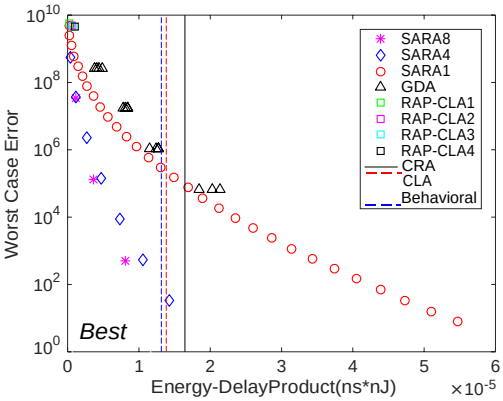


Fig. 13. SARA: The worst case error versus energy-delay product.

sense to additionally compare with static approximate adders, where no configuration is needed. Static approximate adder designs including ETAII [8], FICTS [13] and AFICTS [13] are implemented in the experiment. In addition, CRA-based approximate designs CRA-trunc*i* implemented by ignoring lowest *i* bits in addends are presented, which is a simple but good baseline for comparison. Seven SARA4-DAR designs are obtained based on seven configurations of SARA4 with detection window of 2 bits, while three SARA8-DARare

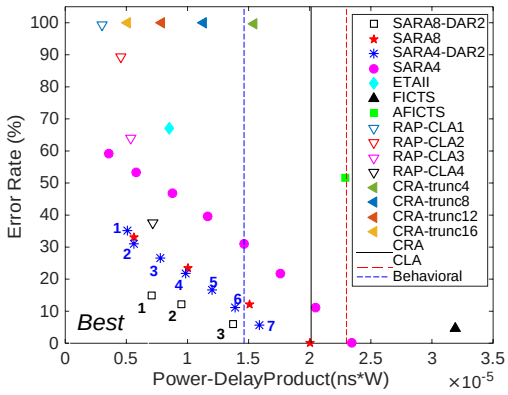


Fig. 14. SARA-DAR: Error rate versus power-delayproduct.

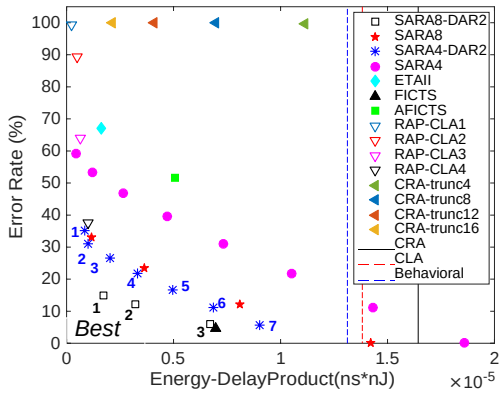


Fig.16.SARA-DAR:Errorrateversusenergy-delayproduct.

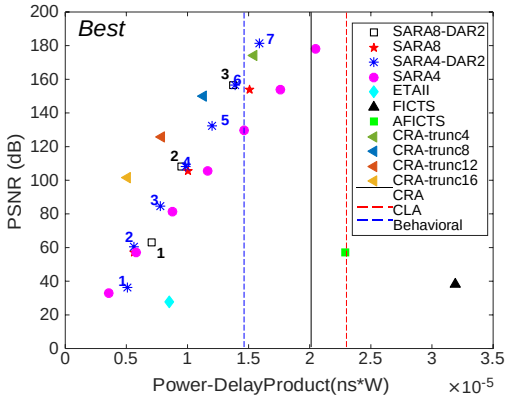


Fig. 15. SARA-DAR: PSNR versus power-delay product.

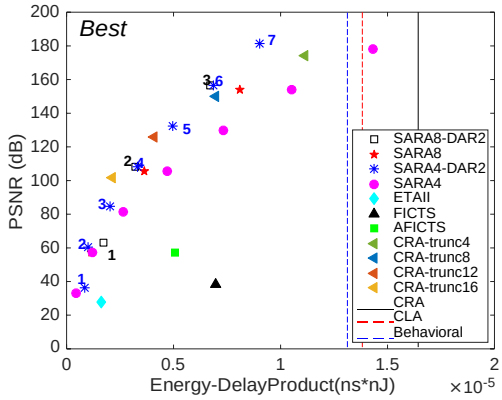


Fig. 17. SARA-DAR: PSNR versus energy-delay product.

basedondifferentconfigurationsofSARA8.Thatis,ifaMUX at bit *i* is configured to accurate carry in SARA4/SARA8, bit *i* of corresponding SARA-DAR is hard-wired to accurate carry without using MUX. When bit *j* in SARA4/SARA8 isin approximation mode, bit *j* of corresponding SARA-DAR uses the delay-adaptivereconfiguration.

Figure 14 shows the error rate versus PDP tradeoff. The dot of SARA4-DAR2 labeled with '1' represents the counterpart of SARA4 when all the MUXes are controlled by delay-adaptive reconfiguration. When we remove the MUX at the highest bit to propagate accurate carry and keep others in delay-adaptive reconfiguration, another SARA4-DAR2 design could be obtained (another dot labeled with '2' in the figure). If we go on to remove more MUXes in MSB, a series of SARA4-DAR2 designs shown as dots with label '3' to '7' can be obtained. Three SARA8-DAR2 designs are created in the same way. According to the figure, the error rate of SARA is mostly lower than RAP-CLA. By using delay-adaptive reconfiguration, SARA-DAR often has less error rate and PDP than SARA. SARA-DAR also greatly outperforms the static approximate adders in both error rate and PDP. Moreover, in Figure 14 we can observe those dots right on the x-axis which represent SARA4 and SARA8 working in accurate mode. Both of them achieve zero error rate. PDP ofaccurateSARA8isabout$2\times10^{-5}ns\cdot W$,whilePDPof accurate SARA4 is almost $2.3\times10^{-5}ns\ W$. In Figure 15, SARA-DARalsodemonstratesbetterPSNR-PDPtradeoffthan other designs, except for comparing with CRA-trunc at some low-PSNRlevels.However,CRA-trunchasalmost100%error rate since it dismiss lower bits in addends, which is the worst among all static approximate adders. Figure 16 and 17 show tradeoff between accuracy and EDP. At a certain EDP level, SARA-DAR has almost the same PSNR as CRA-trunc, which isthebestamongallstaticapproximateadders.

D. Impact of Detection Window in Delay-Adaptive Reconfiguration

Thispartshowstheimpactofdetectionwindowinthetrade- off for delay-adaptive reconfiguration. According to Equa- tion (23), the error rate of MUX output can be reduced by delay-adaptive reconfiguration. As the length of detection window increases, the error rate would decrease because there are less probability that MUX is configured in approximate mode. As a result, the overall error rate varies with the size of detection window. Figure 18 and 19 show the changes of error rate and PSNR of SARA4-DAR with different de-tection window. As the size of detection window increases from 1 to 3, the error rate decreases compared to its SARA counterpart. However, we can observe that the gap of error ratebetweenSARA4-DAR2andSARA4-DAR1varieswith

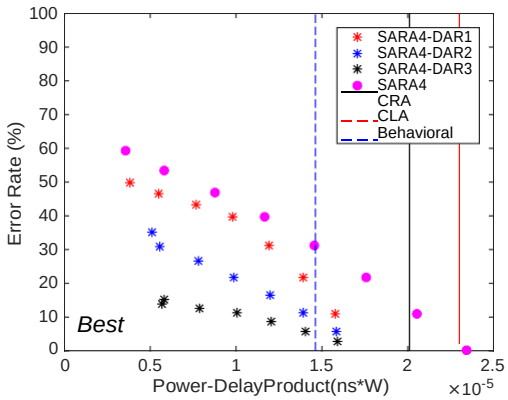


Fig.18. ErrorrateofSARA4-DARwithdifferentdetectionwindow.

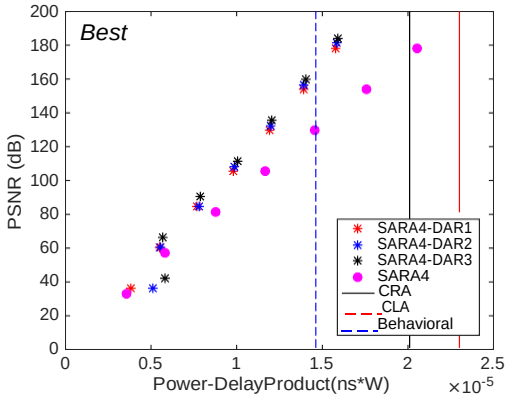


Fig. 19. PSNR of SARA4-DAR with different detection window.

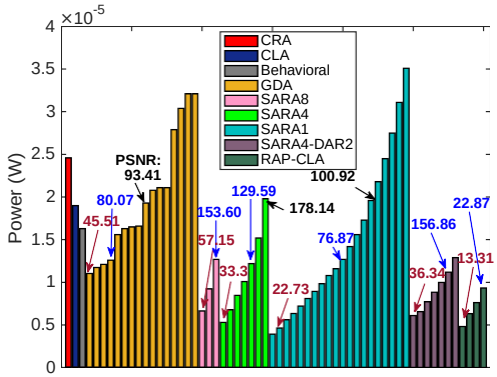


Fig.20.Iso-delaypowercomparison.ThenumbersarePSNR.

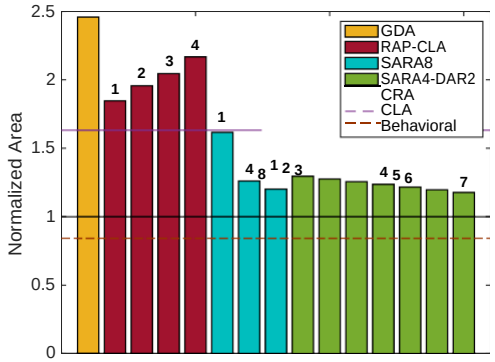


Fig. 21. Area comparison.

different configuration. Although the change in error rate for individual MUX of SARA-DAR is proportional to the size of detection window (as shown in Equation (23)), the overall error rate in output results might not show linear change. When the size of detection window increases by 1, PSNR of SARA4-DAR increases by about 3dB on average. We can also find that the PDP gap between SARA4-DAR2 and SARA4-DAR1 varies with different configurations in both figures. The change of PDP between SARA4-DAR2 and SARA4-DAR1 in most configurations is very small, while it's larger in the first configuration (which is presented as the first dot of SARA4-DAR in the left of the figures). It is mainly attributed to unproportioned change in delay between different configurations.

E. Results of Iso-delay Power and Area

Although power-delay product results have been shown in Sections VI-B and VI-C, the tradeoff between power anddelay is still unclear. The power-delay tradeoff can be obtained by different accuracy configurations or varying supply voltages. Different combinations of configurations and voltages may lead to overwhelming volume of results, which are difficult to interpret, especially when implication to accuracy is involved at the same time. Thus, we indicate the tradeoff by investigatingtheiso-delaypower,whichisthepowerofeach circuit tuned to the same critical path delay ($0.82ns$) by voltage scaling. The results are shown in Figure 20. Ingeneral, SARA4, SARA8 and SARA4-DAR can achieve much lower power than CRA. Although GDA and RAP-CLA seem to provide low power, their PSNR is much less than our designs. Compared at the same iso-delay power level, SARA has more than 20dB increase in PSNR than RAP-CLA, while GDA has more than 70dB decrease than SARA designs. SARA1 shows a large range of iso-power tuning which could reach the lowest and highest power among all adders. We do not have iso-delay power for approximate adders working in accurate mode, because the delay of such case is larger than CRA due to induction of MUXes which cannot provide sufficient room for reducing supplyvoltage.

Last but not the least, we compare area of these designs in Figure 21. Same as our expectation, GDA and RAP-CLA have greater area than CLA while area of SARA4 or SARA8 is significantly smaller than CLA. SARA1 has almost thesame areaasCLAduetoMUXesineverybitwhichaidtheaccuracy configuration. On average, the area of SARA is 39% smaller thanthatofRAP-CLAand50%smallerthanthatofGDA.

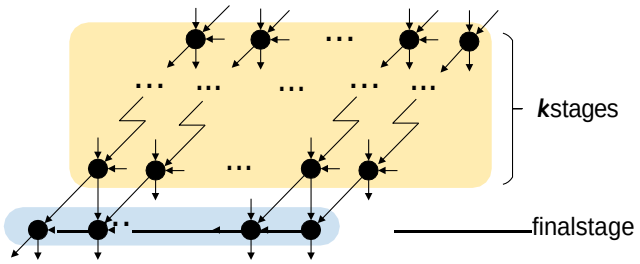


Fig. 22. Basic structure of multiplier.

VII. Applications

A. Extension toMultiplier

In complicated datapath system, multiplier is considered as a much bigger component in power consumption. Our carry-prediction-based approximation uses generate bit to predict the carry from lower segments. The critical delay can be restrainedtoasmallervaluewithshortercriticalpathincarry propagation. Further extension of our technique to multiplier depends on the multiplication structure used in hardware implementation. There is a variety of hardware designs for multiplication, according to the structures of reduction tree. In this section, we apply our technique on three kinds of multiplication structures including array multiplier, Wallace multiplier and Daddamultiplier.

As shown in Figure 22, the basic structure of multiplier employs a three-step process to multiply two integers.

- Step 1: Generate all partial products by using an AND gatearray.
- Step 2: Combine the partial products in *k* stages by layers ofhalf/fulladderuntilthematrixheightisreducedtotwo. Different types of structures depend on the reduction tree used to reduce the number of partial products in thisstep.
- Step 3: Sum the resulting numbers in the final stage by a conventionaladder.

In array multiplier the carry bits in one stage are propagated diagonally downwards, which follows the basic shift-and-add multiplication algorithm. Wallace multiplier based on Wallace tree combines the partial products as early as possible, which makes it faster than array multiplier [20]. Dadda's strategy is to make the combination take place as late as possible, which leads to simpler reduction tree and wider adder in final stage [20]. Thus, we can design approximate multipliers by using ourSARAdesigninsteadofCRAinthefinalstage.

Three types of 16×16 multipliers (array multiplier, Wallace multiplier and Dadda multiplier) as well as behavioral multiplier are synthesized and implemented by using Nangate 45nm Open Cell Library. Their error data are obtained from 100K-run Monte Carlo simulation with uniform distribution of operands. In approximate multiplier the final stage uses SARA4 which consists of sub-adders with bit-width of 4 bits, while the accurate one uses CRA. Figure 23 and 24 present the tradeoff between error and PDP. Most of approximate multipliers configured in approximate mode have better PDP comparedwiththeaccuratemultipliers.Thevarianceoferror between different approximate mode in approximatemultiplier has similar trend as SARA. Total error increases as more bits are configured in approximate mode. Approximate array multiplier shows larger error than approximate Wallace/Dadda multiplier at the same PDP level. It's because array multiplier has larger critical delay from internal stages in step 2 than Wallace/Daddamultiplier.

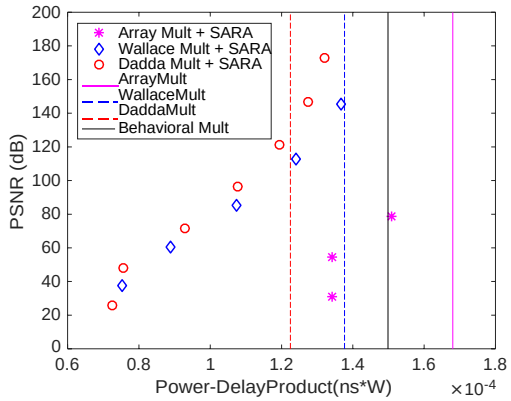


Fig. 23. Multiplier: PSNR versus power-delay product.

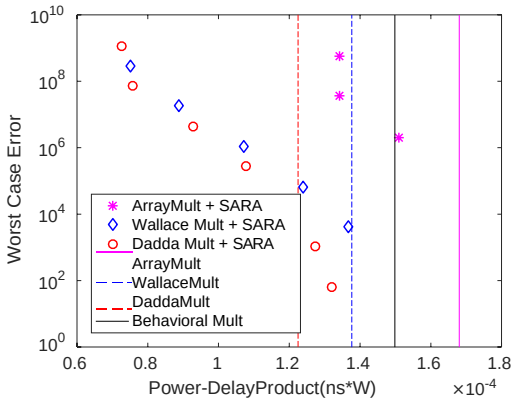


Fig.24. Multiplier:Theworstcaseerrorversuspower-delayproduct.

Figure 25 and 26 show the error versus EDP for both accurate and approximate multipliers. As more MUXes are set to propagate approximate carry, the average error in output increases to about 10^7, which as well achieves best EDP. The worst-case error rate of approximate Dadda multiplier is about 30%, while it comes to about 17% for approximate array multiplier and Wallace multiplier. As shown in Figure 26, when approximate multipliers are working in completely accurate mode (error rate equals 0), EDP is larger than that of their accurate counterpart. In summary, The experimental results show that our technique can be successfully extended to high speed multiplier designs. And due to the simple but effective structure of SARA it provides an easy way for us to convert conventional multiplier into approximatedesign.

B. DCT Computation in ImageProcessing

The discrete cosine transform (DCT) has been recognized asthebasicinmanytransformcodingmethodsforimageand

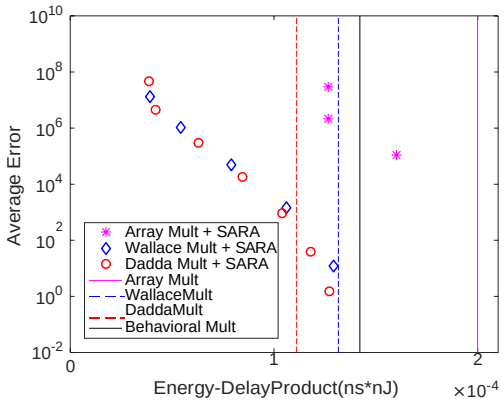


Fig. 25. Multiplier: Average error versus energy-delayproduct.

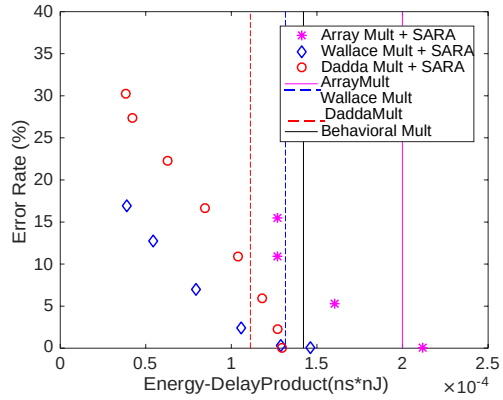


Fig. 26. Multiplier: Error rate versus energy-delay product.

video signal processing. It is used to transform the pixel data of image or video into corresponding coefficients in frequency domain. Since human visual system is more sensitive tothe changes in low frequency, the lost of accuracy in high-frequency components does not heavily degrade the quality of image processed by DCT. In addition, those components in different frequency have different tolerances to thedegradation inoriginaldata.Itisagoodexampletoshowthereconfigurability of our design by applying them in VLSI implementation of DCT computing in JPEG image compression.

The two-dimensional DCT is implemented by the row-column decomposition technique, which contains two stages of 1-D DCT [21]–[23]. The 2-D DCT of size $N \times N$ could be definedas

$$Z = C^t X C \tag{24}$$

where C is a normalized N th-order matrix and X is the data matrix. Generally the image is divided into several $N \times N$ blocks and each block is transformed by 2-D DCT into frequencydomaincomponents.TheVLSIimplementationof DCT computing contains a set of ROM and Accumulator Components(RACs)whichcanbeimplementedbymultipliers andadders[21]–[23].Inthisapplicationweuseapproximate adderstoreplacethoseaccurateonesinRCAstoimplement animprecisebutlowpowercircuitforimageprocessingwhich contains DCTcomputing.

TABLE IV
IMAGE QUALITY COMPARISON IN PSNR

	lenna	cameraman	kiel	house	AVERAGE
Accurate	39.85	38.23	37.68	37.35	38.27
SARA4	38.32	37.50	36.83	36.53	37.30
SARA8	35.33	35.07	34.92	34.81	35.03
SARA4-DAR2	39.45	37.90	37.43	37.00	37.97
GDA	34.53	34.55	34.88	34.20	34.54
RAP-CLA	33.38	33.44	33.51	33.39	33.43

We replace the adders in circuits with different configurations of SARA, SARA-DAR, GDA as well as RAP-CLA.The results are obtained by numerical simulation in Matlab. Aswe know, after DCT process data in different frequency domain have different level of error tolerance. As shown in Figure 31, matrix components in the upper-left corner correspond to lower frequency coefficients which are sensitive to human vision, while those components in lower-right corner might allow moreerrors.

To utilize this feature for better energy-accuracy tradeoff, wemakefollowingconfigurationfordifferentdesigns.

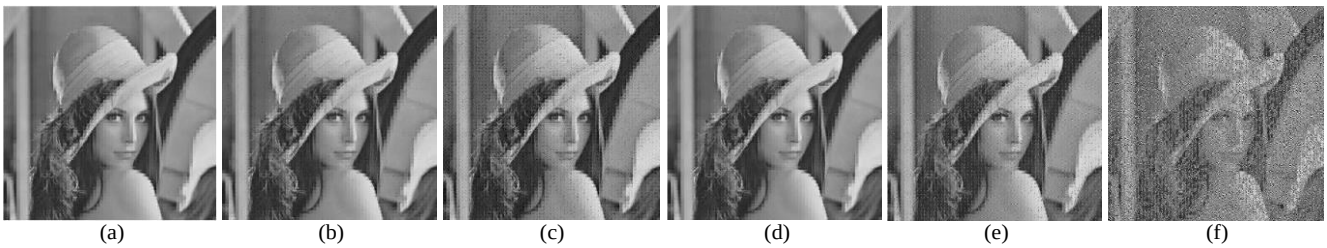
1) **SARA4**: SARA4 with 4, 3, 2, 1 consecutive segments working in accurate mode are used to compute components in S_1, S_2, $S_3$and S_4respectively.
2) **SARA8**: SARA8 with 1 segments in accurate mode are used to compute components in S_1, S_2, while another configuration with all segments in approximate mode are for S_3,S_4.
3) **SARA4-DAR2**: DAR counterpart of SARA4 with detection window of 2bits.
4) **GDA**: $GDA_{4,1}$, $GDA_{3,1}$, $GDA_{2,1}$, $GDA_{1,1}$(same notation as [18]) are used to compute components in S_1, S_2, $S_3$and S_4respectively.
5) **RAP-CLA**: since RAP-CLA can work in one approximate mode, we use RAP-CLA with window size of 20, 16,12,8tocomputecomponentsinS_1,S_2,S_3andS_4.

The image processing results are shown in Table IV. PSNR inthetableisdefinedviathemeansquarederror(MSE).Given an $m \times n$ image I and its restored image K, MSE and PSNR are definedas

$$MSE = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} [I(i,j) - K(i,j)]^2 \tag{25}$$

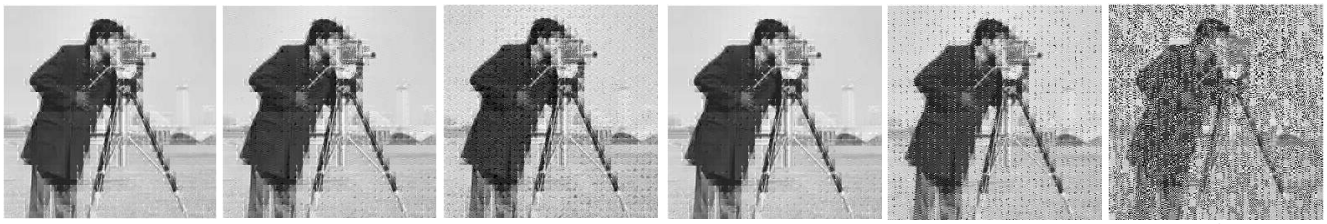
$$PSNR = 20 \cdot \log(MAX_I) - 10 \cdot \log(MSE), \tag{26}$$

where $MAX_I$is the maximum pixel value of the image. SARA4-DAR2 has the highest PSNR for every image among all configurable adders, which is close to the quality of accurate adder. Comparing SARA8 with GDA, they have similar PSNR and similar delay, but SARA8 has less power consumption according to the analysis in the previous section. SARA4-DAR2 achieves better image quality than SARA4, but might result in more power due to additional logics for self-configuration. The image quality for different adders in DCT computing can also be demonstrated in Figure 27 to 30. According to human vision, SARA and its DAR counterpart show better image quality than GDA and RAP-CLA in JPEG compression processing.



(a) (b) (c) (d) (e) (f)

Fig. 27. Comparison of image lenna: (a) accurate adder; (b) SARA4; (c) SARA8; (d) SARA4-DAR2; (e) GDA; (f) RAP-CLA.

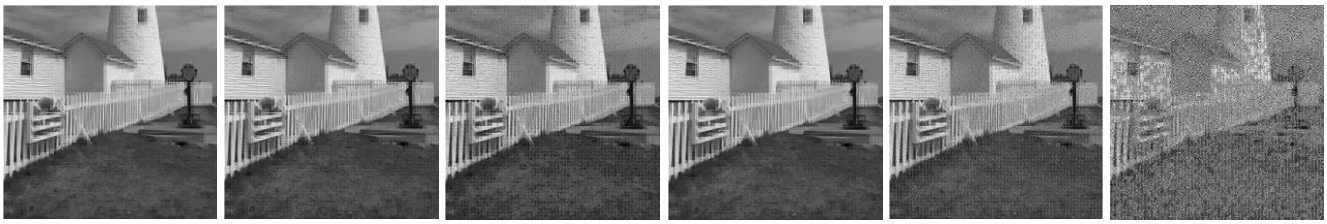


(a) (b) (c) (d) (e) (f)

Fig. 28. Comparison of image cameraman: (a) accurate adder; (b) SARA4; (c) SARA8; (d) SARA4-DAR2; (e) GDA; (f) RAP-CLA.

(a) (b) (c) (d) (e) (f)

Fig. 29. Comparison of image kiel: (a) accurate adder; (b) SARA4; (c) SARA8; (d) SARA4-DAR2; (e) GDA; (f) RAP-CLA.



(a) (b) (c) (d) (e) (f)

Fig. 30. Comparison of image house: (a) accurate adder; (b) SARA4; (c) SARA8; (d) SARA4-DAR2; (e) GDA; (f) RAP-CLA.

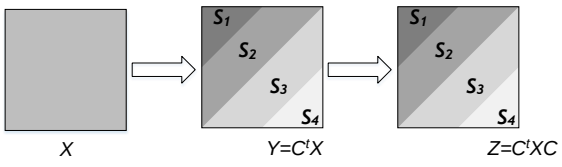


Fig. 31. 2 dimensional descrete cosine transform.

VIII. Conclusion

In this paper, we propse a simple accuracy reconfigurable adder(SARA)design.Ithassignificantlylowerpower/energy-delay product than the latest previous work when comparingat the same accuracy level. In addition, SARA has considerable lower area overhead than almost all the previous works. The accuracy-power-delay efficiency is further improved by a delay-adaptive reconfiguration technique. We demonstrate the efficiency of our adder in the applications of multiplication circuitsandDCTcomputingcircuitsforimageprocessing.

Acknowledgment

The authors would like to thank Dr. Duncan M. Walker from Texas A&M University for helpful discussions.

References

[1]J. Han and M. Orshansky, “Approximate computing: An emerging paradigm for energy-efficient design,” in *IEEE European Test Symposium*, pp. 1–6, 2013.

[2]V. Gupta, D. Mohapatra, A. Raghunathan, and K. Roy, "Low-power digital signal processing using approximate adders," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 32, no. 1, pp. 124–137,2013.
[3]V. Chippa, A. Raghunathan, K. Roy, and S. Chakradhar, "Dynamic effort scaling: Managing the quality-efficiency tradeoff," in *Design Automation Conference (DAC)*, pp. 603–608, 2011.
[4]S.-L. Lu, "Speeding up processing with approximation circuits," *Computer*,vol.37,no.3,pp.67–73,2004.
[5]S. Venkataramani, K. Roy, and A. Raghunathan, "Substitute-and-simplify: A unified design paradigm for approximate and quality configurable circuits," in *Conference on Design, Automation and Test in Europe (DATE)*, pp. 1367–1372,2013.
[6]K. Du, P. Varman, and K. Mohanram, "High performance reliable variable latency carry select addition," in *Conference on Design, Automation and Test in Europe (DATE)*, pp. 1257–1262, 2012.
[7]A. K. Verma, P. Brisk, and P. Ienne, "Variable latency speculative addition: A new paradigm for arithmetic circuit design," in *Conference on Design, Automation and Test in Europe (DATE)*, pp. 1250–1255, 2008.
[8]N. Zhu, W. L. Goh, and K. S. Yeo, "An enhanced low-power high-speed adder for error-tolerant application," in *International Symposium onIntegratedCircuits*,pp.69–72,2009.
[9]N. Zhu, W. L. Goh, W. Zhang, K. S. Yeo, and Z. H. Kong, "Design of low-power high-speed truncation-error-tolerant adder and its application in digital signal processing," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 18, no. 8, pp. 1225–1229, 2010.
[10]N. Zhu, W. L. Goh, and K. S. Yeo, "Ultra low-power high-speed flexible probabilistic adder for error-tolerant applications," in *International SoC Design Conference (ISOCC)*, pp. 393–396, 2011.
[11]G. Liu, Y. Tao, M. Tan, and Z. Zhang, "Casa: correlation-aware speculative adders," in *International Symposium on Low Power Electronics and Design*, pp. 189–194, 2014.
[12]J. Hu and W. Qian, "A new approximate adder with low relative error and correct sign calculation," in *Conference on Design, Automation and Test in Europe (DATE)*, pp. 1449–1454,2015.
[13]J. Miao, K. He, A. Gerstlauer, and M. Orshansky, "Modeling and synthesis of quality-energy optimal approximate adders," in *International Conference on Computer-Aided Design (ICCAD)*, pp. 728–735, 2012.
[14]S. Mazahir, O. Hasan, R. Hafiz, M. Shafique, and J. Henkel, "An area-efficient consolidated configurable error correction for approximate hardware accelerators," in *Design Automation Conference (DAC)*, pp. 1–6, 2016.
[15]A. B. Kahng and S. Kang, "Accuracy-configurable adder for approximate arithmetic designs," in *Design Automation Conference (DAC)*, pp. 820–825, 2012.
[16]M. Shafique, W. Ahmad, R. Hafiz, and J. Henkel, "A low latency generic accuracy configurable adder," in *Design Automation Conference (DAC)*, pp. 1–6, 2015.
[17]V. Benara and S. Purini, "Accurus: A fast convergence technique for accuracy configurable approximate adder circuits," in *IEEE Computer Society Annual Symposium on VLSI*, pp. 577–582, 2016.
[18]R. Ye, T. Wang, F. Yuan, R. Kumar, and Q. Xu, "On reconfiguration-oriented approximate adder design and its application," in *International Conference on Computer-Aided Design (ICCAD)*, pp. 48–54, 2013.
[19]O. Akbari, M. Kamal, A. Afzali-Kusha, and M. Pedram, "RAP-CLA: A reconfigurable approximate carry look-ahead adder," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. PP, no. 99, pp. 1–1, 2017.
[20]W. J. Townsend, E. E. Swartzlander, and J. A. Abraham, "A comparison of dadda and wallace multiplier delays," *Proc. SPIE, Advanced Signal Processing Algorithms, Architectures, and Implementations*, vol. 13, pp. 552–560,2003.
[21]S. Yu and E. Swartziander, "DCT implementation with distributed arithmetic," *IEEE Transactions on Computers*, vol. 50, no. 9, pp. 985–991, 2001.
[22]M.-T. Sun, T.-C. Chen, and A. M. Gottlieb, "VLSI implementation of a 16* 16 discrete cosine transform," *IEEE Transactions on Circuits and Cystems*, vol. 36, no. 4, pp. 610–617, 1989.
[23]D. Gong, Y. He, and Z. Cao, "New cost-effective VLSI implementation of a 2-d discrete cosine transform and its inverse," *IEEE Transactions on Circuits and Systems for VideoTechnology*, vol. 14, no. 4, pp. 405–415.

Wenbin Xu received the B.S. and M.S. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2008 and 2011 respectively. He is now working toward the Ph.D. degree in computer engineering at Texas A&M University, College Station, TX. He is currently a Research Assistant in the Department of Electrical and Computer Engineering, Texas A&M University. His research interests include computer-aided design, approximate computing and hardwaresecurity.

Sachin Sapatnekar (S'86, M'93, F'03) receivedthe B. Tech. degree from the Indian Institute of Technology, Bombay, the M.S. degree from Syracuse University, and the Ph.D. degree from the University of Illinois. He taught at Iowa State University from 1992 to 1997 and has been at the University of Minnesota since 1997, where he holds the Distinguished McKnight University Professorship and the Robert and Marjorie Henle Chair. He has received seven conference Best Paper awards, a Best Poster Award, two ICCAD 10-year Retrospective Most Influential Paper Awards, the SRC Technical Excellence award and the SIA University Researcher Award. He is a Fellow of the ACM and the IEEE.

Jiang Hu (F'16) received the B.S. degree in optical engineering from Zhejiang University (China) in 1990, the M.S. degree in physics in 1997 and the Ph.D. degree in electrical engineering from the University of Minnesota in 2001. He worked with IBM Microelectronics from January 2001 to June 2002. In 2002, he joined the electrical engineering faculty at Texas A&M University. His research interest is in optimization for large scale computing systems, especially VLSI circuit optimization, adaptive design, hardware security, resource allocation andpower management in computing systems.

Dr.HureceivedtheBestPaperAwardattheACM/IEEEDesignAutomation Conference in 2001, an IBM Invention Achievement Award in 2003, and the Best Paper Award from the IEEE/ACM International Conference on Computer-Aided Design in 2011. He has served as a technical program committee member for DAC, ICCAD, ISPD, ISQED, ICCD, DATE, ISCAS, ASP-DAC and ISLPED. He is general chair for the 2012 ACM International Symposium on Physical Design, and an associate editor of IEEE transactionson CAD 2016-2011. Currently he is an associate editor for the ACM transactions on Design Automation of Electronic Systems. He received a Humboldt research fellowship in2012.

Carbon Nanotube Using Electronics

1Kukatla Rajeshwar 2Singamaneni Nagajyothi

Abstract

We evaluate the potential of carbon nanotubes (CNTs) as the basis for a new nanoelectronic technology. After briefly reviewing the electronic structure and transport properties of CNTs, we discuss the fabrication of CNT field-effect transistors (CNTFETs) formed from individual single-walled nanotubes (SWCNTs), SWCNT bundles, or multiwalled (MW) CNTs. The performance characteristics of the CNTFETs are discussed and compared to those of corresponding silicon devices. We show that CNTFETs are very competitive with state-of-the-art conventional devices. We also discuss the switching mechanism of CNTFETs and show that it involves the modulation by the gate field of Schottky barriers at the metal-CNT junctions. This switching mechanism can account for the observed subthreshold and vertical scaling behavior of CNTFETs, as well as their sensitivity to atmospheric oxygen. The potential for integration of CNT devices is demonstrated by fabricating a logic gate along a single nanotube molecule. Finally, we discuss our efforts to grow CNTs locally and selectively, and a method is presented for growing oriented SWCNTs without the involvement of a metal catalyst.

Keywords*—*Carbon nanotubes (CNTs), field-effect transistors (FETs), molecular electronics, nanoelectronics.

I. INTRODUCTION

Carbon nanotubes (CNTs) are hollow cylinders composed of one or more concentric layers of carbon atoms in a honeycomb lattice arrangement. Multiwalled nanotubes (MWCNTs) were observed for the first time in transmission electron microscopy (TEM) studies by Iijima in 1991 [1], while single-walled nanotubes (SWCNTs) were produced independently by Iijima [2] and Bethune [3] in 1993.

SWCNTs typically have a diameter of 1–2 nm and a length of several micrometers. The large aspect ratio makes the nanotubes nearly ideal one-dimensional (1-D) objects, and as such the SWCNTs are expected to have all the unique properties predicted for these low-dimensional structures [4]–[7]. In addition, as we discuss below, depending on the detailed arrangement of the carbon atoms the SWCNTs can be metallic or semiconducting [8], [9]. Furthermore, the C–C bonds in CNTs are very strong, resulting in an extremely high mechanical stability (Young's modulus about ten times higher than that of steel) and chemical inertness. The strong, covalent bonding also leads to near perfect side-wall structures with very few defects.

CNTs are currently considered as promising building blocks of a future nanoelectronic technology. This is not simply due to their small size but rather to their overall properties. In fact, many of the problems that silicon technology is or will be facing are not present in CNTs. Below we list some of these CNT properties and their implications for electronics.

1) Carrier transport is 1-D. This implies a reduced phase space for scattering of the carriers and opens up the possibility of ballistic transport. Correspondingly, power dissipation is low. Furthermore, as we discuss in Section II, their electrostatic behavior is different from that of silicon devices with implications on screening and electron/hole tunneling.
2) All chemical bonds of the C atoms are satisfied and there is no need for chemical passivation of dangling bonds as in silicon. This implies that CNT electronics would not be bound to use SiO_2 as an insulator. High dielectric constant and crystalline insulators can be used, allowing, among other things, the fabrication of three-dimensional (3-D) structures.
3) The strong covalent bonding gives the CNTs high mechanical and thermal stability and resistance to electromigration. Current densities as high as 10^9 A/cm^2 can be sustained [10].
4) Their key dimension, their diameter, is controlled by chemistry, not conventional fabrication.
5) In principle, both active devices (transistors) and interconnects can be made out of semiconducting and metallic nanotubes, respectively.

We see that the properties of the SWCNTs are truly remarkable. However, finding ways to effectively exploit these properties remains a challenge. In the rest of this paper

Manuscript received October 2, 2002.

The authors are with the IBM Research Division, T. J. Watson Research Center, Yorktown Heights, NY 10598 USA (e-mail: avouris@us.ibm.com).

Digital Object Identifier 10.1109/JPROC.2003.818338

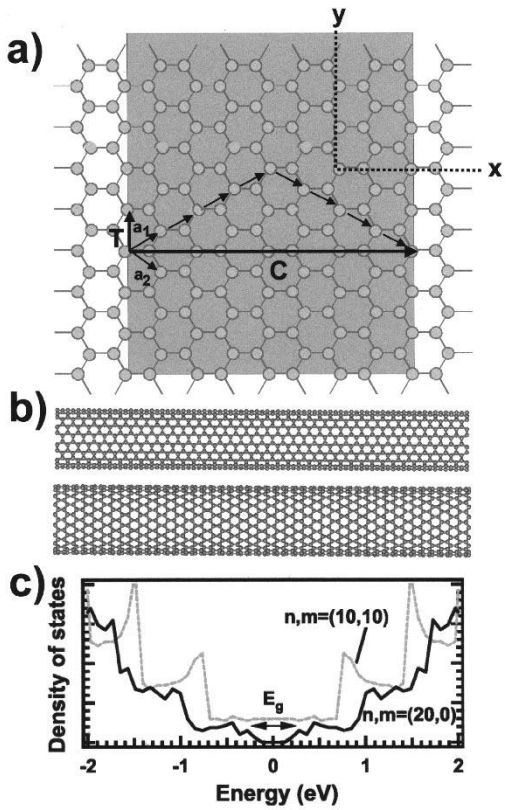


Fig. 1. (a) Description of the structure of CNTs in terms of the chirality vector C and the (n,m) indices. The example shown involves a (4,4) tube. (b) Atomic structure of a metallic (10,10) CNT (top) and a semiconducting (20,0) CNT (bottom), and (c) the corresponding density of states versus energy plots.

we will review the electronic structure of the CNTs and then present our efforts to fabricate nanotube transistors and simple integrated circuits and understand the underlying device physics.

II. Electronic Structure of Carbon Nanotubes

The electronic structure and electrical properties of SWCNTs are usually discussed in terms of the electronic structure of a graphene sheet (a layer of graphite) [4]–[9]. The SWCNT can be thought of as being formed by folding a piece of graphene to give a seamless cylinder. The circumference of the nanotube is expressed by the so-called chirality vector, $\mathbf{C}$, connecting two crystallographically equivalent sites of the two-dimensional (2-D) graphene sheet (see Fig. 1). $\mathbf{C} = n\mathbf{a}_1 + m\mathbf{a}_2$, where $\mathbf{a}_1$ and $\mathbf{a}_2$ are the unit vectors of the hexagonal honeycomb lattice, so that any nanotube can be described by a pair of integers (n,m) that define its chiral vector. For example, the chiral vector shown in Fig. 1(a) describes a (4,4) nanotube. The unit shell of the nanotube is defined as the rectangle formed by $\mathbf{C}$ and the 1-D translational vector $\mathbf{T}$ identified in Fig. 1. Also shown are the atomic structures of a (10,10) and a (20,0) CNT [Fig. 1(b)] and their density of states [Fig. 1(c)].

The interesting electrical properties of CNTs are due in a large part to the peculiar electronic structure of the graphene. Its band structure and its hexagonal first Brillouin zone

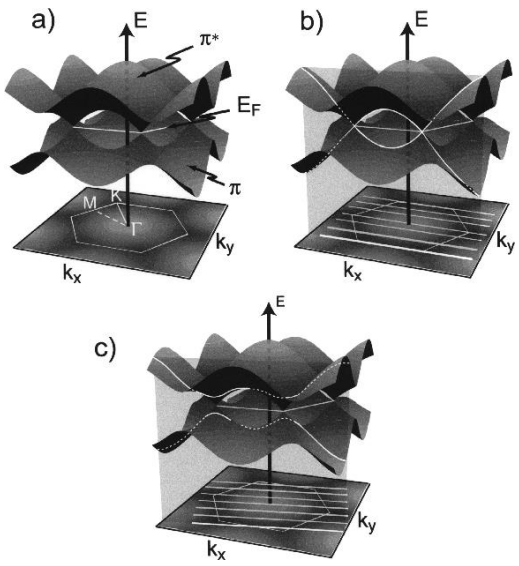


Fig. 2. (a). Band structure of a graphene sheet (top) and the first Brillouin zone (bottom). (b) Band structure of a metallic (3,3) CNT. (c) Band structure of a (4,2) semiconducting CNT. The allowed states in the nanotubes are cuts of the graphene bands indicated by the white lines. If the cut passes through a K point, the CNT is metallic; otherwise, the CNT is semiconducting.

are shown in Fig. 2(a). The energy surfaces describing the valence (π) and conduction (π^*) states touch at six points (Fermi points) lying at the Fermi level. This unusual band structure has immediate consequences for the electronic properties of graphene. While allowed states exist at the Fermi level, the dimensionality of the system (2-D) results in a vanishing density of states when integrating over the Fermi surface. Because of this particular situation, graphene is a zero-gap semiconductor.

In the case of a nanotube there is an additional quantization arising from the confinement of the electrons in the circumferential direction in the tube. This requires that the circumferential component of the wave vector $\boldsymbol{k}_C$ can only take the values fulfilling the condition $\boldsymbol{k}_C \cdot \mathbf{C} = 2\pi j$ where $\mathbf{C}$ is again the chirality vector and j an integer. As a result, each band of graphene splits into a number of 1-D subbands labeled by j. Fig. 2(b) shows the states of a (3,3) CNT. The allowed energy states of the tube are cuts of the graphene band structure. When these cuts pass through a Fermi point, as in the case of the (3,3) nanotube, the tube is metallic. In cases where no cut passes through a K point, the tubes are semiconducting [Fig. 2(c)]. It can be shown that an (n,m) CNT is metallic when $n = m$, it has a small gap (due to curvature-induced $\sigma - \pi$ mixing) when $n - m = 3i$ [11], where i is an integer, while CNTs with $n - m \neq 3i$ are truly semiconducting [8], [9].

As long as we restrict our interest to low energies (i.e., a few hundred meV from the Fermi energy E_F) the band structure of a metallic nanotube can be approximated by two sets of bands with a linear dispersion intersecting at k_F and $-k_F$ [see white lines in Fig. 2(b)]. Electrons with $dE/dk > 0$ move to the right, while electrons with $dE/dk < 0$ move to the left. In semiconducting CNTs the two bands do not cross at E_F, but a diameter-dependent band gap develops

with $E_{GAP} = (4\hbar v_F/3d_{CNT})$, where d_{CNT} is the tube's diameter and v_F the Fermi velocity [8], [9]. The above theoretical predictions have been confirmed experimentally by scanning tunneling spectroscopy [12], [13].

The SWCNTs are 1-D objects and as such their two-terminal conductance is given by Landauer's equation: [14], [15] $G = (2e^2/h) \sum_i^N T_i$ where $2e^2/h$ is the quantum of conductance and T_i is the transmission of a contributing conduction channel (subband). The sum involves all contributing conduction channels, i.e., channels whose energy lies between the electrochemical potentials of the left and right reservoirs to which the nanotube is connected. In the absence of any scattering, i.e., when all $T_i = 1$, the resistance $(R = 1/G)$ of a metallic SWCNT is $h/(4e^2) \approx 6.5$ kΩ because, as we discussed above, $N = 2$. This quantum mechanical resistance is a contact resistance arising from the mismatch of the number of conduction channels in the CNT and the macroscopic metal leads.

There is strong evidence that in the case of metallic SWCNTs, so that these tubes behave as ballistic conductors [16]–[19]. This arises from the 1-D confinement of the electrons which allows motion in only two directions. This constraint along with the requirements for energy and momentum conservation severely reduces the phase space for scattering processes. However, in addition to the quantum mechanical contact resistance, there are other sources of contact resistance, such as those produced by the existence of metal-nanotube interface barriers, or poor coupling between the CNT and the leads. These types of resistance are very important and can dominate electrical transport in nanotubes, especially at low temperatures where typically they lead to charging and the observation of Coulomb blockade phenomena. Localization can also be induced by contacts to metallic electrodes, a fact that makes four-probe measurements very difficult, requiring special arrangements [19].

Unlike SWCNTs, the electrical properties of MWCNTs have received less attention. This is due to their complex structure; every carbon shell can have different electronic character and chirality, and the presence of shell–shell interactions [20], [21]. However, at low bias and temperatures, and when MWCNTs are side-bonded to metallic electrodes, transport is dominated by outer-shell conduction [10], [22], [23]. MWCNTs show 1-D or 2-D characteristics, depending on their diameter and the property considered.

III. Fabrication and Performance of Carbon Nanotube FETs

FETs, particularly in CMOS form, have been proven to be the most technologically useful device structures. It is, thus, natural that we have chosen to build such devices using CNTs. The first such devices were fabricated in 1998 [24], [25]. In these a single SWCNT was used to bridge two noble metal electrodes prefabricated by lithography on an oxidized silicon wafer as shown in Fig. 3. The SWCNT played the role of the "channel," while the two metal electrodes functioned as the "source" and "drain" electrodes. The heavily doped silicon wafer itself was used as the "gate" (back-gate). These CNTFETs behaved as p-type FETs (we will return to this point in Section V) and had an I(on)/I(off) current ratio of $\approx 10^5$. While functional, the devices had a high parasitic contact resistance (≥ 1MΩ), low drive currents (a few nanoamperes), low transconductance $g_m \sim 1$ nS, and high inverse subthreshold slopes S ~1–2 V/decade. To a large extent the unsatisfactory characteristics were due to bad contacts. The CNT was simply laid on the gold electrodes and was held by weak van der Waals forces. To improve the contacts we adapted a different fabrication scheme where the semiconducting SWCNTs (s-SWCNTs) were dispersed on an oxidized Si wafer, and the source and drain electrodes, now made of metals that are compatible with silicon technology such as Ti or Co, were fabricated on top of them [26]. Thermal annealing of the contacts, which in the case of Ti electrodes led to the formation of TiC, produced a stronger coupling between the metal and the nanotube and a reduction of the contact resistance [26], [27].

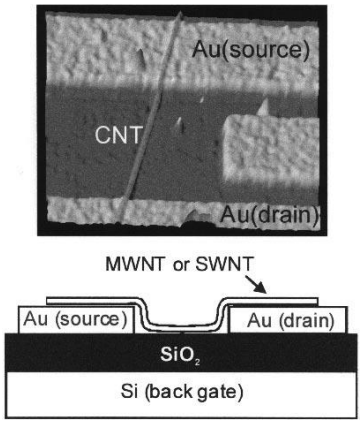


Fig. 3. Top: AFM image of one of our early CNTFETs. Bottom: Schematic cross section of the CNTFET [25].

Fig. 4(a) and (b) shows the output and transfer characteristics of such a CNTFET with Co electrodes [26]. From the $I_d - V_{gs}$ curves we see that the transistor is p-type and has a high on–off current ratio of $\sim 10^6$. This new CNTFET configuration has a significantly reduced contact resistance, ~ 30 kΩ, a much higher current in the μA range, and a transconductance $g_m = 0.34\mu$S, i.e., ~ 200 times higher than that of van der Waals-bonded CNTFETs.

All of the early devices were back-gated with very thick gate insulators (SiO_2 thickness $t_{ox} \approx$ 100-150 nm). As with conventional MOSFETs we should be able to improve their performance by increasing the gate capacitance by reducing the insulator thickness or increasing the dielectric constant. However, unlike in the MOSFET configuration where the capacitance is similar to that of a plane capacitor, i.e., gate capacitance $C_G \sim 1/t_{ox}$, the CNTFET geometry will predict a $C_G \sim \ln^{-1}(at_{ox} + b)$ dependence [25]. In addition to increasing the gate capacitance, it is essential that each CNTFET is gated independently by its own gate so that complex integrated circuits can be built.

A next generation of CNTFETs with top gates was fabricated by dispersing SWCNTs on an oxidized wafer. Atomic

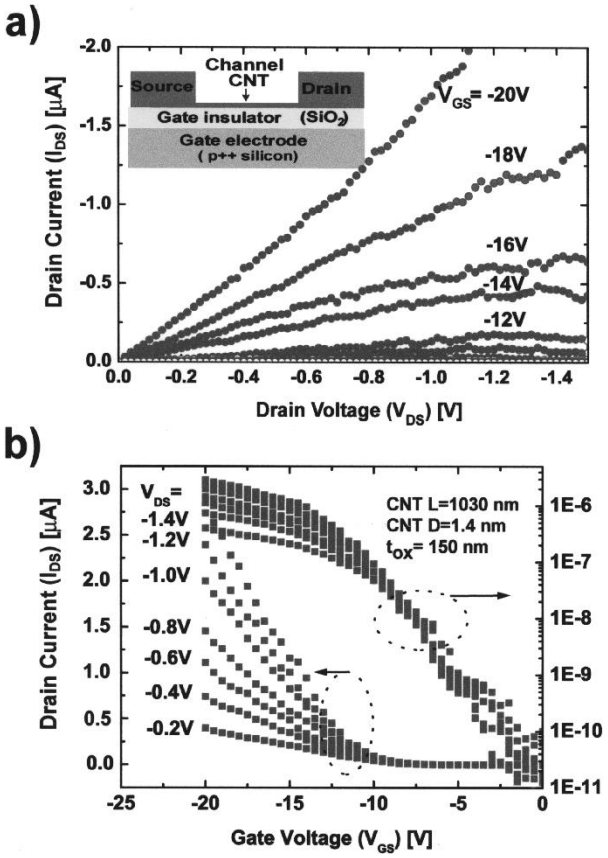


Fig. 4. (a) Output characteristics of a CNTFET with cobalt source and drain electrodes deposited on top of tube. (b) The corresponding transfer characteristics.

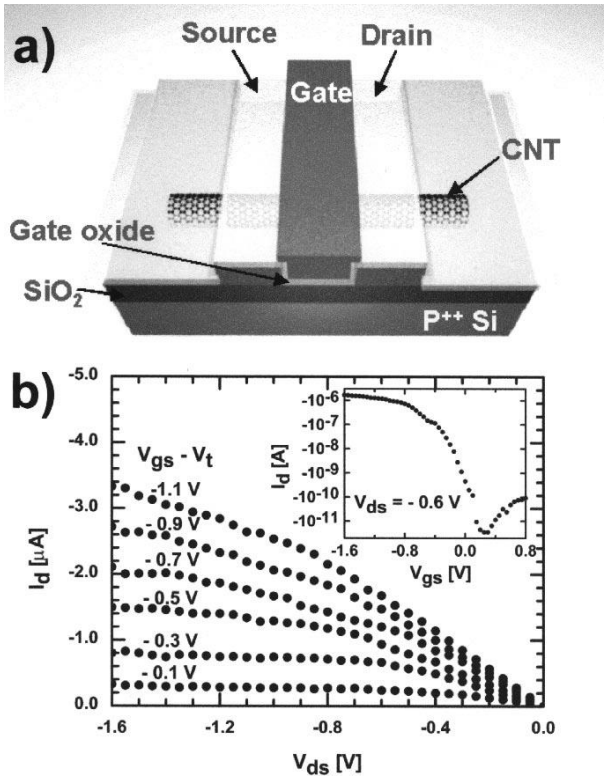


Fig. 5. (a) Schematic representation of one of our top-gated CNTFET with Ti source, drain, and gate electrodes. A 15-nm SiO_2 film was used as the gate oxide. (b) The *I–V* characteristics of the device.

force microscopy (AFM) imaging was used to identify single CNTs, and the Ti source and drain electrodes were fabricated on top of by e-beam lithography and liftoff [28]. After annealing at 850 °C to transform the contacts into TiC [27], a 15- to 20-nm-gate dielectric film was deposited by chemical vapor deposition (CVD) from a mixture of SiH_4 and O_2 at 300 °C. After annealing for ~0.5 h at 600 °C in N to $_2$ densify the oxide, 50-nm-thick Ti or Al gate electrodes were patterned by lithography and liftoff.

In Fig. 5(a) we show a schematic of a top-gated CNTFET, and in Fig. 5(b) the output characteristics of such a device with Ti electrodes and a 15 nm SiO_2 gate insulator film [28]. Such a CNTFET can also be switched by the bottom gate (wafer) and the resulting characteristics can be compared with those of the device under top-gate operation. This device has a superior performance; the threshold voltage of the top-gated CNTFET is significantly lower, –0.5 V, than under bottom-gated operation, +2 V. Similarly, the drive current is much higher under top gating, and the transconductance is similarly high, $g_m = 3.3$ μS per nanotube.

Since the eventual objective of nanotube electronics is to be competitive with silicon electronics, it is important to compare their relative performances, despite the fact that the CNTFETs are still far from being optimized. In these experiments a single SWCNT is used, so we express the current carrying capabilities of the devices per unit width (per micrometer) as is the practice in microelectronics. The

Table 1
Comparison of Key Performance Parameters for a 260 nm-Long Top Gate p-CNTFET, a 15-nm-Bulk Si p-MOSFET, and a 50- nm SOI p-MOSFET

	p-type CNTFET	Ref a	Ref. b
Gate Length (nm)	260	15	50
Gate oxide thickness (nm)	15	1.4	1.5
V_t (V)	-0.5	~-0.1	~-0.2
I_{ON} (μA/μm) ($V_{DS} = V_{GS} = V_t - 1$ V)	2,100	265	650
I_{OFF} (nA/μm)	-150	<-500	-9
Subthreshold slope (mV/dec)	130	~100	70
Transconductance (μS/μm)	2,321	975	650

a) B. Yu et al. IEDM Tech. Dig. 2001, p. 934; b) R. Chau et al. IEDM Tech. Dig. 2001, p. 621

diameter of the s-SWCNTs used was ~1.4 nm. Table 1 shows the comparison of the characteristics of the CNTFET with two recent high-performance Si p-channel devices: a 15-nm-gate length MOSFET built on bulk silicon [29] and a 50-nm-gate length device built using SOI technology [30]. We also note that an optimal device layout may require the use of an array of CNTFETs. In this case the results can be scaled to give the total current for the array [26]. The current would increase the denser the packing of the tubes. However, screening at close separations can reduce the actual current per CNT by up to a factor of 2 [31].

From Table 1 we see that the CNTFET is capable of delivering three to four times higher drive currents than the Si MOSFETs at an overdrive of 1 V, and has about four times higher transconductance. From the above and other considerations, it is clear that CNTFETs, even in this early stage of de-

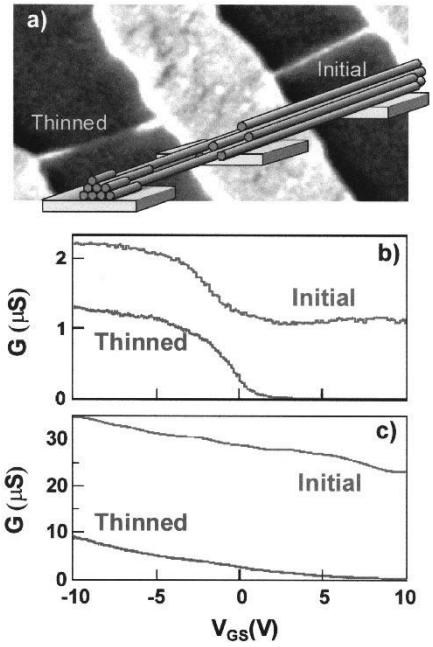


Fig. 6. Conductance (G) as a function of gate voltage (v_{GS}) of a CNT bundle containing both metallic and semiconducting nanotubes before and after selective breakdown of the metallic CNTs. (a) Images of the intact and thinned nanotube bundle. (b) G versus v_{GS} for a thin bundle. (c) The same for a very thick bundle. In the latter case some semiconducting tubes had to be sacrificed in order to remove the innermost metallic tubes in the bundle.

velopment, can be very competitive with the corresponding Si devices. Further refinements can be expected by additional reductions in t_{ox} and the use of high-ε insulators. We have already seen significant improvements by using HfO_2 as a gate insulator [32], [52]. An insight to the ultimate potential of CNTFETs is provided by a recent study [33]. In this nanotube FET, the role of the gate was played by a droplet of an electrolyte connected to an electrochemical electrode. The combination of an electrolyte dielectric constant of about 80 and of the ultrathin (~0.5 nm) Hemholtz layer in the electrolyte led to an extremely high transconductance of about 20 μS. Further opportunities for improvements in the CNTFET performance arise from new insights on the switching mechanism in the CNTFET, as will be discussed in Section V.

IV. Transistors From Nanotube Bundles: Bundle Composition Engineering

A major impediment to the large-scale fabrication of CNTFETs is the fact that the current synthetic schemes for SWCNTs generate mixtures of metallic (m) and semiconducting (s) nanotubes. These tubes tend to adhere to each other, forming "bundles," or "ropes" [34]. No good methods exist for the preparation of only m- or s-CNTs by selective synthesis or postsynthesis separation. If CNTFETs were to be fabricated from such a bundle, the m-CNTs in the bundle would short out the device, as shown in Fig. 6. Currently, dilute suspensions of CNT bundles are ultrasonicated and dispersed on a wafer. AFM imaging is then used to identify isolated CNTs to build the CNTFET. We have developed

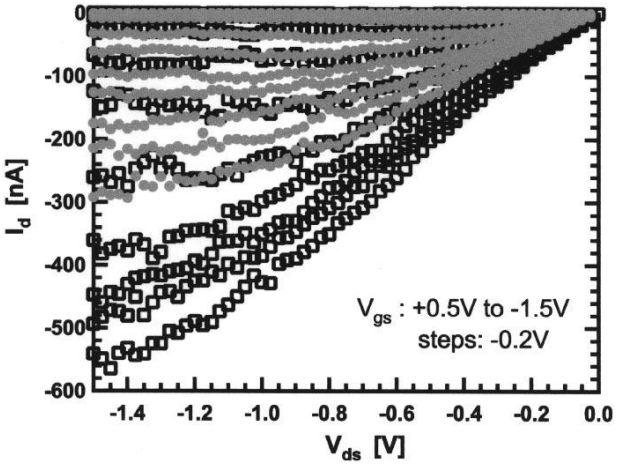


Fig. 7. Drain current versus gate voltage curves of a CNTFET upon interchange of its source and drain [32], [52].

the technique of "constructive destruction" [20] that allows us to selectively destroy the m-CNTs in a rope, leaving the s-CNTs intact.

As we discussed earlier, CNTs can carry enormous current densities at low electron energies. At higher energies, however, optical phonon excitation is possible [35], [36]. This leads to current saturation and the deposition of large amounts of energy, which eventually destroys the CNT structure [36]. To apply this method to remove m-CNTs from bundles, we first deposit the bundles on an oxidized Si wafer, then we fabricate on them an array of source drain and side gate electrodes. By applying an appropriate voltage bias to the gate, the s-CNTs can be depleted of their carriers. Then when a sufficiently high source-drain bias V_{ds} is applied, the generated current passes only through the m-CNTs, leading to their destruction, while leaving the s-CNTs essentially intact (see Fig. 6) [20]. In this way arrays of CNTFETs can be generated [20].

V. The Switching Mechanism of Carbon Nanotube Transistors

Up to this point we have implicitly assumed that the CNTFET switching mechanism is the same as that of conventional silicon devices. However, a number of observations suggest otherwise. For example, Fig. 7 shows the I_d versus V_{ds} curves for the same CNTFET upon interchange of the source and drain [32], [52]. A different current is obtained in the two cases. If the operation of the device were to be dictated by the properties of the bulk CNT, then the saturation current would be the same, since both sets of curves are taken with the same CNT. However, the different characteristics can be accounted for if transport in the tube is dominated by barriers (Schottky barriers[1]) at the CNT–metal contacts. In that case a small asymmetry of the barriers at the source and drain junctions could result in different saturation

[1]We use the term Schottky barriers in its most general sense to account for band bending in a semiconductor at a metal/semiconductor interface. When arguing about the switching in CNT transistors, we assume that there is no additional barrier present in our devices.

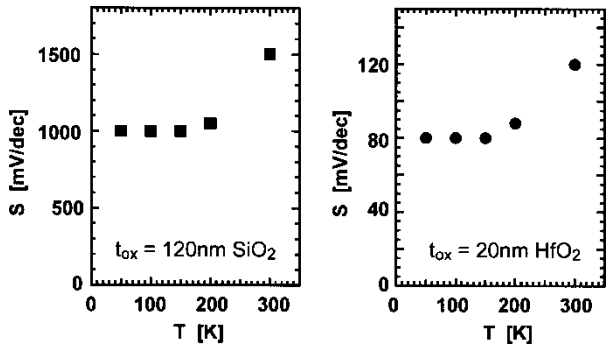


Fig. 8. Temperature dependence of the inverse subthreshold slopes, S, of two CNTFETs with 120 nm of SiO_2 and 20 nm of HfO_2 gate oxides, respectively [32], [52].

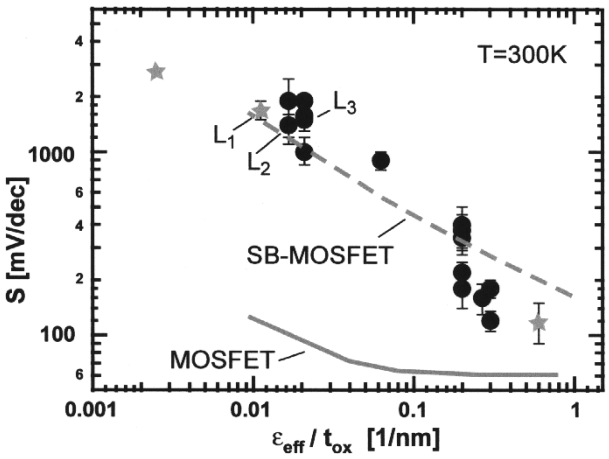


Fig. 9. Dependence of the inverse subthreshold slope S on the ratio ε_{eff}/t_{ox}, where ε_{eff} is the effective dielectric constant and t_{ox} is the thickness of the gate oxide. The points are experimental data from our own work, as well as from the literature. The solid line is the predicted behavior for bulk switching, while the dashed line is the prediction for the Schottky barrier switching model [32], [52].

currents. Further support for the existence of these barriers comes from the study of the subthreshold characteristics of CNTFETs.

In long-channel FET devices, the drain current I_D varies exponentially with V_G, and for drain bias $V_D > 3\,k_B T/q$ it is essentially independent of V_D [37]. A device characteristic of particular importance is the gate-voltage swing, or inverse subthreshold slope S. This is given by $S = \ln 10 \cdot dV_G/d(\ln I_D) \simeq (k_B T/q) \cdot \ln 10(1 + C_D/C_G)$. For a fully depleted device, the depletion capacitance C_D is zero; therefore, the second term in parenthesis becomes one. CNTs are a perfect example of a device exhibiting $C_D = 0$, since no charge variation can occur across the tube circumference. Under these ideal conditions S depends only on the temperature and has a value of 60 mV/dec at 300 K. Deviations appear when interface trap states are present in the oxide [37]. The capacitance due to these interface states $C_{\rm int}$ is in parallel with the depletion capacitance. The early back-gated CNTFETs with thick (100–150 nm) gate oxides had unexpectantly high S values of 1–2 V/dec. Devices with thin oxides, such as the top-gated CNTFET in Fig. 5, have S$\sim$100–150 mV/dec. Interface traps could in principle account for these observations. However, we found that the high S values for thick oxides are very similar independent of whether an n-type, a p-type, or an ambipolar transistor was measured. Since there is no doubt about the presence of Schottky barriers in case of an ambipolar device, and—as we proved [32], [52]—a Schottky barrier model alone can explain the trend of S as a function of $t_{\rm ox}$ quantitatively, interface traps are not responsible for the high S values in case of thick gate oxides. Further evidence that S is determined by Schottky barriers comes from its temperature dependence. As can be seen from Fig. 8, S is temperature dependent at higher temperatures but levels off at temperatures below about 200 K, suggesting a carrier tunneling process. The described subthreshold behavior may be unexpected for a bulk-switched device, but can be fully accounted for by a Schottky barrier transistor model [32], [52]. Calculations of S as a function of $t_{\rm ox}$ based on the SB and bulk models along with our own and literature experimental data are shown in Fig. 9. The data are fitted well by the SB model but not by the bulk switching model.

Focusing on the Schottky barrier itself, we note that the electrostatics in 1-D is different than that in 3-D and it is reflected in the shape of the barriers [38]. In 1-D the barriers are thin, showing an initial sharp drop followed by a logarithmically decreasing tail. Because of the thinness of the barrier, tunneling in 1-D is easy and can dominate transport. In Fig. 10(a) we show the schematic of a top-gated CNTFET and the electric field generated by the gate in that structure. Fig. 10(b) gives the conductance of the device as a function of the gate voltage for different values of the Schottky barrier height. Finally, Fig. 10(c) shows the source Schottky barrier for a mid-gap CNT at three different gate voltages [38]. The shape of the barrier and the thinning resulting from the gate field can be clearly seen.

The SB model can also provide an explanation for the long-standing problem involving the effect of the ambient on the performance of CNTFETs. Already in the first studies of CNTFETs, it was observed that although the CNTs used were not intentionally doped, the fabricated CNTFETs were p-type [24], [25]. These early FETs were fabricated in air. It was later found that when CNTs are placed in vacuum, their electrical resistance increased and their thermopower changed sign [40]. It was proposed that the CNTs transfer electrons to atmospheric O_2 and, thus, become doped with holes [41].

We have performed detailed studies of this gas effect on nanotubes in a CNTFET configuration [42]. In our early studies, we used a back-gated FET configuration with a thick (100–150 nm) gate oxide. As Fig. 11 shows, initially the air-exposed CNTFET was p-type. However, after annealing under vacuum becomes n-type. Furthermore, as Fig. 11 shows this unexpected transformation is reversible; reexposure to O_2 brings back the p character of the FET. Intermediate stages where the CNTFET exemplifies *ambipolar* behavior are clearly seen.

The above findings prove that the p character of the CNTFET is not an intrinsic property of the CNTs, but results

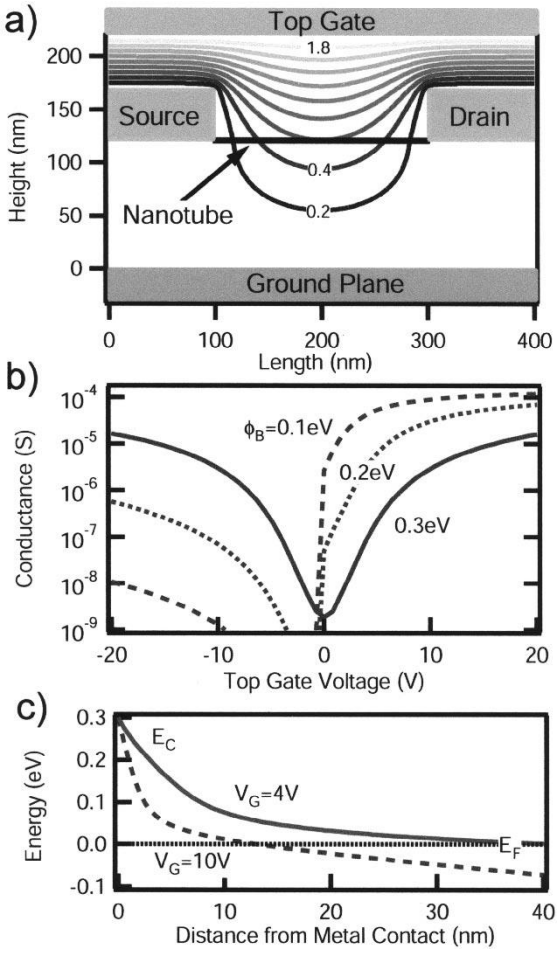


Fig. 10. Simulation of the operation of a top-gated Schottky barrier CNTFET. (a) The structure of the CNTFET and the electric field lines ($v_{gate} = 2$ V).(b) Conductance versus gate voltage for different values of the Schottky barrier. (c) The effect of gate bias on the Schottky barrier at the source–CNT junction [39].

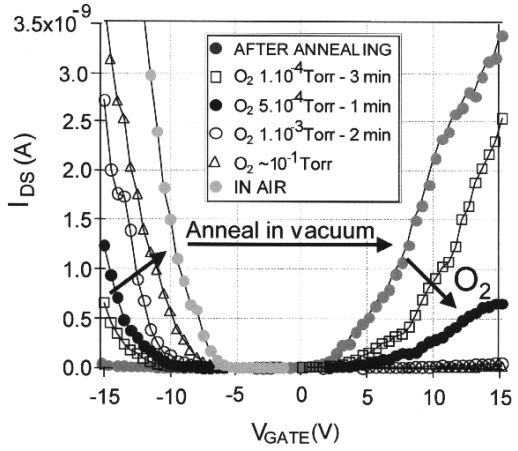


Fig. 11. Transformation of a p-type CNTFET by annealing in vacuum into an n-type CNTFET and the reverse transformation upon exposure to O_2. The transformation proceeds via ambipolar intermediate states of the CNTFET. No threshold shifts are observed, and the drain current at $v_{GATE} = 0$ does not change.

from the interaction with O_2. We have performed a number of experiments to ascertain the nature of this interaction. In Fig. 12 we show the behavior of a CNTFET upon doping

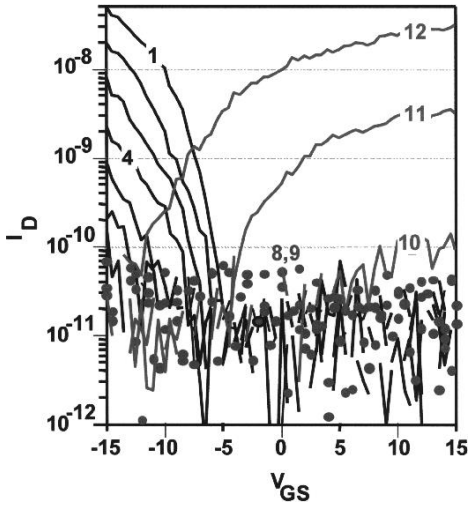


Fig. 12. The effect of doping with increasing amounts of potassium on the electrical characteristics of an initially p-type (curve 1) CNTFET.

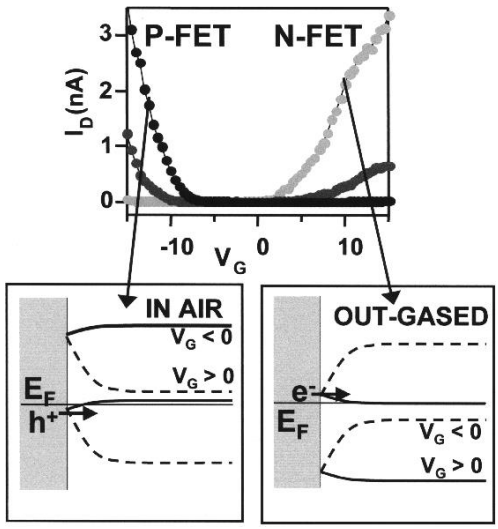


Fig. 13. Qualitative diagram showing the lineup of the valence and conduction bands of a CNT with the metal Fermi level at the source–CNT junction first in air and after annealing in vacuum.

with an electron donor, in this case potassium [42]. The well-known characteristics of doping, i.e., a shift of the threshold gate voltage and an increasing current at $V_g = 0$, are clearly seen. However, this behavior is in stark contrast with the behavior observed upon O_2 exposure (see Fig. 11). From this and other experiments, we have concluded that although some doping by O_2 may take place, this by itself cannot account for the observed behavior.

As we have already shown above, transport and switching in CNTFETs is controlled by the Schottky barriers at the contacts. Oxygen then must affect these barriers. When a nanotube is bonded to a metal electrode the resulting charge transfer determines the lineup of the nanotube bands. However, this charge transfer and the resulting field can be strongly affected by the coadsorption of other species such as oxygen near or at the CNT–metal junction. These coadsorbates can change locally the surface potential [39] or

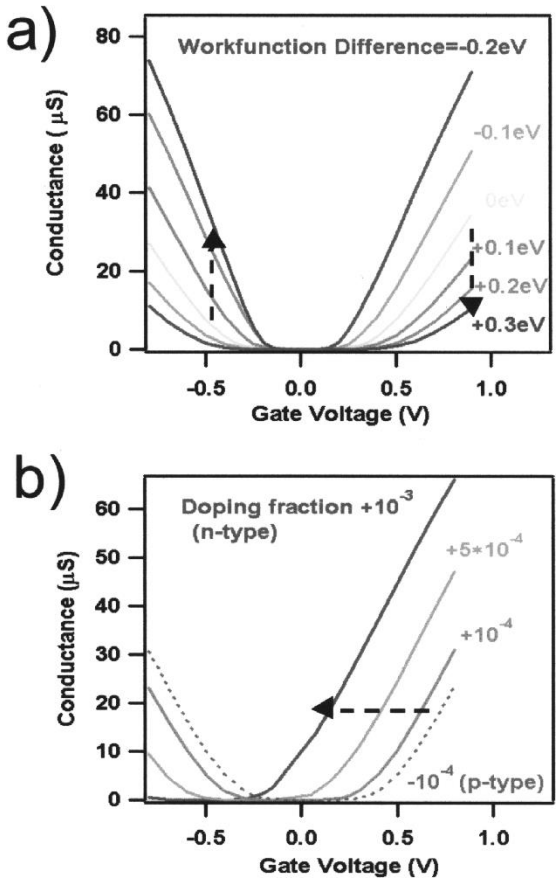


Fig. 14. Simulations of the effect of O and K on the CNTFETs current–voltage characteristics [39].

directly interact with the junction [43]. In Fig. 13 we give a schematic that accounts qualitatively for the band lineup at a metal–CNT junction in air and after annealing in a vacuum. The p character in air is the result of Fermi level pinning near the valence band maximum.

Support for this interpretation is provided by theoretical modeling. In order to take advantage of the simpler electrostatics, we used a model of a CNT surrounded by a cylindrical gate. In Fig. 14 we simulate the I–V characteristics of a CNTFET upon oxygen adsorption by changing the local surface potential [39]. Although the numbers cannot be compared because of the different device geometry used in the calculation, it is clear that the resulting behavior is similar to that of Fig. 11. In contrast, a model where fixed charges are placed along the length of the tube in order to simulate a doping interaction gives a behavior similar to that of doping by potassium in Fig. 12.

The effect of oxygen dominates the behavior of CNTFETs as long as the gate field is weak. This was the case for our earlier structures with thick gate oxides. However, in our recent CNTFETs with thin oxides ($t_{\rm ox}$ ~2–5 nm), we observe ambipolar characteristics even in air, indicating a near midgap lineup of the CNT bands [53]

VI. Multiwalled Nanotube Field-Effect Transistors

Our discussion above has been limited to transistors made out of SWCNTs. Low-temperature studies of the

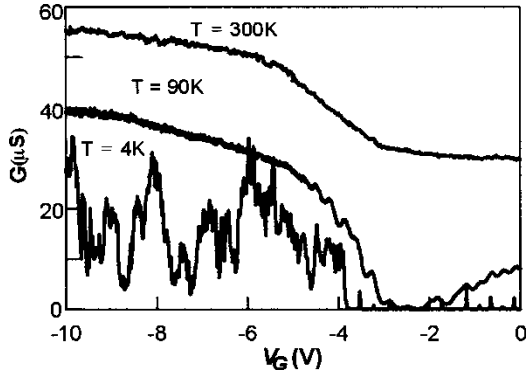


Fig. 15. Temperature dependence of the conductance G of an MWCNT as a function of the gate voltage.

Aharonov–Bohm effect in MWCNTs have concluded that in MWCNTs side-bonded to metal electrodes, effectively only the outer shell contributes to electrical transport [22]. One would expect then that MWCNTs with a semiconducting outer shell could be used to fabricate CNTFETs. However, in semiconducting CNTs, the band gap ($E_{\rm Gap}$) is inversely proportional to the tube diameter; therefore, only small-diameter MWCNTs are expected to display large switching ratios at room temperature.

In Fig. 15 we show the effect of the gate voltage V_G on the conductance G of a 14-nm diameter MWCNT, at different temperatures [21]. Clearly, the gate can modulate the conductance of the MWCNT channel at room temperature. However, there is a large residual conductance, which can be assigned to the coupling of the outer semiconducting shell to an inner metallic shell. This coupling is expected to be activated with an activation energy of the order of $\sim E_{\rm Gap}/2$. At low temperatures, the contribution of the inner metallic shell is suppressed (see Fig. 15). These observations indicate that the transport characteristics of an MWCNT at ambient temperature may have contributions from more than the outer shell even for side-bonded CNTs. Because of the activated nature of the shell-to-shell transport, and given that the semiconducting gap is inversely proportional to the CNT diameter, multishell transport should be more important for large-diameter tubes.

When the applied $V_{\rm DS}$ is increased, the average energy of the carriers is expected to increase, as well as the coupling between the carbon shells. As we discussed in Section IV, higher electron energies lead to energy dissipation and breakdown. Our studies of MWCNTs have shown that initiation of the breakdown occurs at a power threshold, which is significantly lower in air than in vacuum [36]. This indicates that an "oxidation" process takes place in air. The oxidation is most likely not a purely thermal process. Calculations on graphite have shown that once certain defects are generated, e.g., di-vacancies, a self-sustaining chain reaction with oxygen can take place [44]. Because of the lower breakdown threshold in air, the shell breakdown in MWCNT proceeds sequentially from the outer to the inner shells. This shell-by-shell breakdown can be seen as a stepwise decrease in the current flowing through the CNT [see Fig. 16(a)]. It is also illustrated by AFM images of an MWCNT draped over

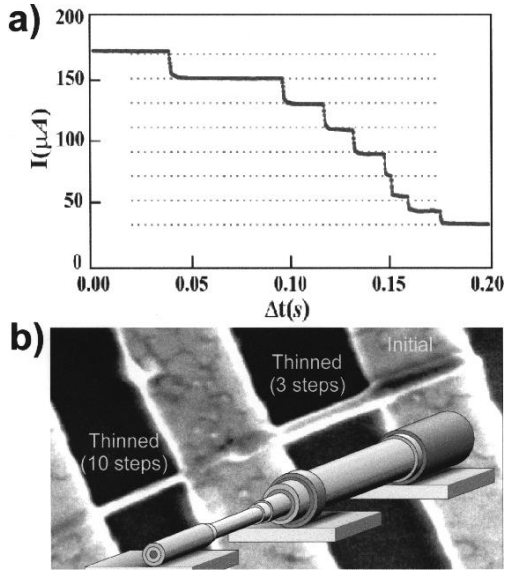


Fig. 16. (a) Stepwise decrease of the current during the breakdown of an MWCNT. (b) Images of an MWCNT showing part of the initial tube, as well as segments of the same MWCNT from which three and ten carbon atom shells have been removed.

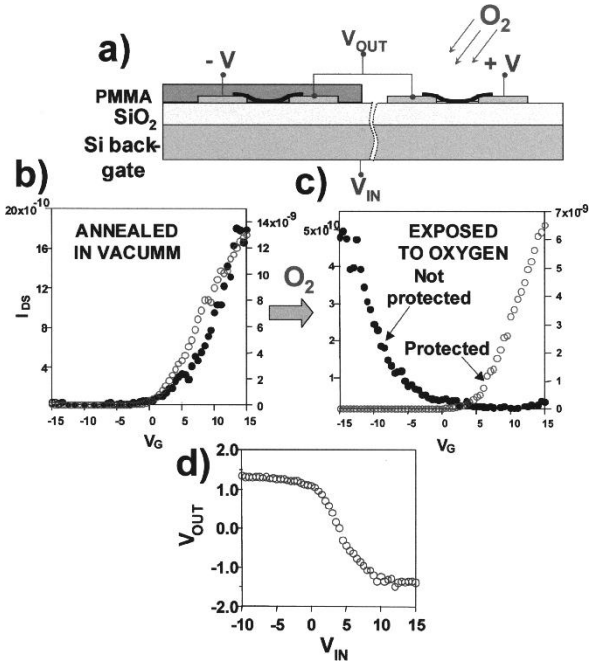


Fig. 17. (a) Schematic representation of the internanotube voltage inverter (NOT gate). One of the FETs is protected by a layer of PMMA. (b) After annealing two p-CNTFETs in vacuum to form two n-type CNTFETs. (c) After exposure to oxygen at 300 K. (d) Electrical behavior of the inverter.

several metal electrodes, so that individual CNT segments can be electrically stressed independently. Thinned segments that have lost three and ten shells, respectively, as well as a portion of the initial MWCNT, are clearly seen [Fig. 16(b)]. The capability to remove carbon atom shells one by one and identify their character through the effect of the gate field [20], [21], along with the diameter dependence of the band gap ($E_{\mathrm{Gap}} \sim 1/d_{\mathrm{CNT}}$) allows the fabrication of CNTFETs with a desired E_{Gap} can be fabricated using the controlled breakdown process.

VII. Carbon Nanotube Integrated Circuits: Logic Gates

So far we have concentrated our discussion on the performance of individual CNTFETs. The fabrication of integrated circuits using such devices is the next step. In 2001 we demonstrated that this was possible by fabricating a CMOS-like voltage inverter (a logic NOT gate) [45]. For simplicity, we used the early design of CNTFETs involving the CNT on top of gold electrodes

In Fig. 17 we show the structure and electrical characteristics of an inverter circuit involving a n- and a p-CNTFETs. Originally, both CNTFETs were p-type because of their exposure to air (oxygen). We then covered one of them by a protective film of PMMA (a more stable protection is provided by a SiO_2 film [27]), while the other was left unprotected. Both of the CNTFETs were then annealed under vacuum, which transformed both of them into n-type. After cooling, the pair was exposed to oxygen, which converted the unprotected CNTFET to p-type, while the protected one remained n-type. In this way, the two complementary CNTFETs needed were formed and wired.

The inverter works the same way as an ordinary CMOS inverter. The input voltage is applied simultaneously to the gates of the complementary CNFETs. The p-CNFET is polarized by a positive voltage, the n-FET by a negative voltage, and a common contact is used as the intermolecular inverter's output. A positive input voltage turns the n-CNFET "on" (the p-CNFET being "off"), resulting in the transmission of the negative voltage to the output. A negative input, on the other hand, turns the p-CNFET "on," producing a positive output. The electrical characteristics of the CNTFETs at each stage and those of the resulting inverter are shown in Fig. 17. We refer to this inverter circuit as an *intermolecular* inverter because it involves two nanotube molecules.

Ideally, one would like to achieve the ultimate level of integration by fabricating the circuits along the length of a single CNT, i.e., form an *intramolecular* circuit. A first realization of this approach is shown in Fig. 18(a) [44]. As the AFM image shows, the nanotube is placed on top of three prefabricated gold electrodes. In this way two back-gated initially p-type CNTFETs are formed. Then they are both covered by PMMA and a window is opened by e-beam lithography over the channel of one of them. Through this window, the channel is n-doped using potassium as a dopant. In this way two complementary FETs are formed along the same nanotube. The electrical characteristics of the resulting inverter are shown in Fig. 18(b). It is particularly interesting that despite the fact that no effort was made to optimize the construction and performance of the individual CNTFETs, the resulting inverter had a gain of almost two. This suggests that optimized CNTFETs would lead to much higher gain and can be wired along the length of a single CNT to produce more complex circuits. Following this initial work [45], more nanotube logic gates of complementary [46] or transistor–resistor [47] type have been demonstrated.

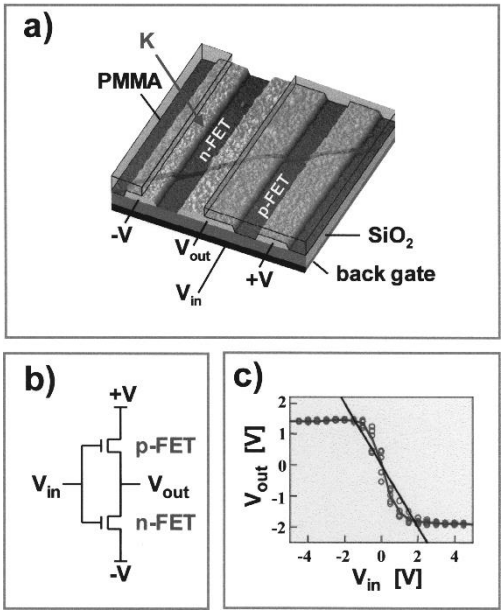


Fig. 18. (a) Atomic force microscope image of the intrananotube voltage inverter. (b) Schematic of the inverter. (c) Electrical behavior of the intrananotube inverter.

VIII. Nanotube Synthesis

SWCNTs are produced using arc discharges [2], [3], laser ablation of a carbon target [34], or CVD [48]. In all of these techniques, a metal catalyst (typically Co, Fe, and/or Ni) in the form of nanoparticles is utilized.

Currently, catalytic CVD is the most widely used technique [48]. One of the advantages of this approach is that it allows nanotubes to be grown locally by placing the catalyst at the appropriate location [49]. We have experimented with catalytic CVD using a combination of electrolytic and lithographic approaches to control both the location and orientation of the growth. The procedure shown in Fig. 19 involves the following steps:

1) patterning a thin, heavily doped silicon surface layer lithographically;
2) electrolytically etching the silicon to form porous silicon on the sidewalls of the patterned structure;
3) protecting with photoresist the areas of the silicon surface where we do not want CNT growth;
4) driving into the exposed pores the metal catalyst, and after removing the rest of the photoresist, reacting with CH_4 at 1000 C to form SWCNTs.

Fig. 20 shows SEM images of SWCNTs grown this way bridging adjacent silicon pads. These pads can subsequently be metallized.

Quite often, the presence of heavy metal catalyst particles in the nanotube product is unwanted but their removal is problematic and usually leads to damaged nanotubes. In applications in nanoelectronics, it is also likely that arrays of parallel oriented tubes will be needed in order to reduce the impedance of the devices and provide a high drive current [26]. Oriented growth of tubes is a very desirable way to achieve this type of nanotube organization. Postsynthesis alignment has also been pursued [50].

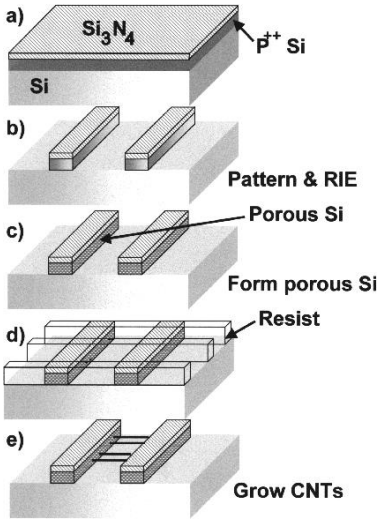


Fig. 19. Sequence of lithographic and other processing steps used to produce selective local growth of SWCNTs by chemical vapor deposition using CH_4 as the source of carbon.

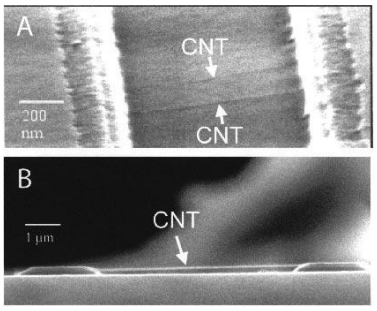


Fig. 20. Electron microscope images (top and side views) showing locally grown CNTs connecting silicon electrodes.

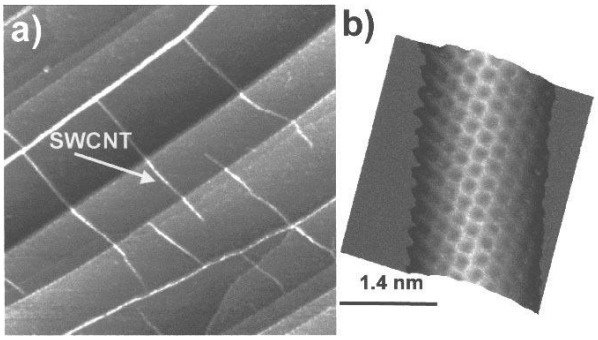


Fig. 21. (a) STM image of oriented (parallel) SWCNTs produced by heating under high vacuum to 1650 C a 6 H-SiC wafer with a (0001) Si-face. (b) Atomic resolution STM image of a SWCNT produced by this method.

Recently, we discovered a way to produce oriented SWCNTs by a catalyst-free approach [51]. This approach involves the thermal annealing under vacuum of SiC crystals. Specifically, 6H–SiC wafers with a polished (0001) Si face (this surface is equivalent to the (111) surface of cubic SiC (3 C–SiC)) were heated to 1650 C at 10^{-9} torr.

Fig. 21(a) shows a scanning tunneling microscope (STM) image of a sample cut along the (1, 1,0,0) axis of the wafer resulting in a morphology characterized by well-ordered

parallel steps in the (1,1,–2,0) direction. The white 1-D structures, identified as the CNTs, extend over several steps or terraces and are not present on the samples annealed in vacuum at a temperature below 1400°C. An atomic resolution image of a semiconducting nanotube is shown in Fig. 21(b). From such images, as well as TEM images, the diameter of the nanotubes produced was determined to be in the range of 1.2–1.6 nm. These tubes are seen to have their axis perpendicular to the surface steps or be aligned along the steps. Extensive STM and AFM studies showed that this organization is uniform over their entire area of the sample. When the surface morphology is composed of terraces, the nanotubes form a weblike network with a predominance of 120°angles between straight sections (not shown) [51].

Along with individual SWCNTs, thicker tubes are also observed. Manipulation experiments using the tip of an AFM [51] indicate that these structures are SWCNT bundles, not MWCNTs.

The orientation of the nanotubes could be a result of the growth process or may involve a postsynthesis rearrangement. Our AFM experiments show that the tubes are mobile on the surface at the high temperature used for their formation. This is deduced by the observation that after perturbing the structure of the nanotubes by AFM manipulation, annealing at a temperature of 1300°C, i.e., at a temperature lower than that needed for nanotube formation, brings the manipulated nanotubes back to their parallel orientation. Thus, we believe that the orientation of the nanotubes results from their motion that releases part of the mechanical stress incorporated in the randomly grown network by favoring straight segments and by matching their orientation to the crystallographic structure of the surface. At the same time, formation of bundles contributes to the lowering of the total energy.

The above findings suggest that: 1) it may be possible to orient preformed nanotubes on an inert substrate by heating them at a temperature at which they acquire sufficient mobility and 2) in principle, it is possible to synthesize nanotubes in a controlled manner by patterning graphene strips followed by annealing. By selecting the direction of the cut of the strip, the chirality of the resulting nanotube can be chosen.

IX. Conclusion and the Future

CNTs are new materials with outstanding electrical properties. The high conductivity and exceptional stability of metallic nanotubes makes them excellent candidates for future use as interconnects in nanodevices and circuits. FETs using semiconducting CNTs have operating characteristics that are as good as or better than state-of-the-art silicon devices, and significant improvements should be expected in the near future. However, while CNTs are one of the most promising materials for molecular electronics, many challenges remain before they can become a successful technology. Most challenging are the materials issues. We still lack a method that produces a single type of CNT. In this respect, seeded growth techniques are a possibility and need to be explored. Another possible solution involves the development of efficient separation techniques, and work is pursued in this direction with encouraging initial results. The sensitivity of the electrical properties of CNTs and CNT devices to the nature of the CNT–metal contacts and the ambient environment demonstrated in this article shows that better understanding and control of these problems is absolutely essential. For CNT device integration, new fabrication techniques that are based on self-assembly of CNTs are highly desirable. While our own current interest is in computer electronics, it is likely that the initial applications of CNT devices will be in less integrated systems such as sensors, or in special applications where devices of exceptional miniaturization and performance are needed. Apart from their technological importance, CNTs are ideal model systems for the study and understanding of transport in 1-D systems and for the development of molecular fabrication technologies.

References

[1] S. Iijima, "Helical microtubules of graphitic carbon," *Nature*, vol. 354, pp. 56–58, 1991.

[2] S. Iijima and T. Ichihashi, "Single-shell carbon nanotubes of 1 nm diameter," *Nature*, vol. 363, pp. 603–605, 1993.

[3] D. S. Bethune, C. H. Kiang, M. S. Devries, G. Gorman, R. Savoy, J. Vaszquez, and R. Beyers, "Cobalt-catalyzed growth of carbon nanotubes with single atomic layer walls," *Nature*, vol. 363, pp. 605–607, 1993.

[4] M. S. Dresselhaus, G. Dresselhaus, and P. Avouris, Eds., *Carbon Nanotubes: Synthesis, Structure Properties and Applications*. Berlin, Germany: Springer-Verlag, 2001.

[5] P. L. McEuen, M. S. Fuhrer, and H. Park, "Single-walled carbon nanotube electronics," *IEEE Trans. Nanotechnol.*, vol. 1, pp. 78–85, Mar. 2002.

[6] C. Dekker, "Carbon nanotubes as molecular quantum wires," *Phys. Today*, vol. 52, p. 22, 1999.

[7] P. G. Collins and P. Avouris, "Nanotubes for electronics," *Sci. Amer.*, vol. 283, pp. 38–45, 2000.

[8] M. S. Dresselhaus, G. Dresselhaus, and R. Saito, "Carbon fibers based on C_{60} and their symmetry," *Phys. Rev. B*, vol. 45, pp. 6234–6242, 1992.

[9] J. W. Mintmire, B. I. Dunlap, and C. T. White, "Are fullerene tubules metallic?," *Phys. Rev. Lett.*, vol. 68, pp. 631–634, 1992.

[10] S. Frank, P. Poncharal, Z. L. Wang, and W. A. deHeer, "Carbon nanotube quantum resistors," *Science*, vol. 280, pp. 1744–1746, 1998.

[11] M. Ouyang, J.-L. Huang, C. L. Cheung, and C. M. Lieber, "Energy gaps in 'metallic' single-walled carbon nanotubes," *Science*, vol. 292, pp. 702–705, 2001.

[12] J. W. Wildoer, L. C. Venema, A. G. Rinzler, R. E. Smalley, and C. Dekker, "Electronic structure of atomically-resolved carbon nanotubes," *Nature*, vol. 391, pp. 59–62, 2001.

[13] T. W. Odom, J. L. Huang, P. Kim, and C. M. Lieber, "Atomic structure and electronic properties of single-walled nanotubes," *Nature*, vol. 391, pp. 62–64, 2001.

[14] Y. Imry and R. Landauer, "Conductance viewed as transmission," *Rev. Mod. Phys.*, vol. 71, pp. S306–S312, 1999.

[15] S. Data, *Electronic Transport in Mesoscopic Systems*. Cambridge, U.K.: Cambridge Univ. Press, 1995.

[16] P. L. McEuen, M. Bockrath, D. H. Cobden, Y.-G. Yoon, and S. G. Louie, "Disorder, pseudospins, and backscattering in carbon nanotubes," *Phys. Rev. Lett.*, vol. 83, pp. 5098–5101, 1999.
[17] L. Wenjie, M. Bockrath, D. Bozovic, J. H. Hafner, M. Tinkman, and P. Hongkum, "Fabry–Perot interference in a nanotube electron waveguide," *Nature*, vol. 411, pp. 665–669, 2001.
[18] J. Kong, E. Yenilmez, T. W. Tombler, W. Kim, H. Dai, R. B. Laughlin, L. Liu, C. S. Jayanthi, and S. Y. Wu, "Quantum interference and ballistic transmission in nanotube electron waveguides," *Phys. Rev. Lett.*, vol. 87, pp. 106 801/1–106 891/4, 2001.
[19] J. Appenzeller, R. Martel, P. Avouris, H. Stahl, and B. Lengeler, "Optimized contact configuration for the study of transport phenomena in ropes of single-wall carbon nanotubes," *Appl. Phys. Lett.*, vol. 78, pp. 3313–3315, 2001.
[20] P. G. Collins, M. S. Arnold, and P. Avouris, "Engineering carbon nanotubes using electrical breakdown," *Science*, vol. 292, pp. 706–709, 2001.
[21] P. G. Collins and P. Avouris, "Multishell conduction in multiwalled carbon nanotubes," *Appl. Phys. A*, vol. 74, pp. 329–332, 2002.
[22] A. Bachtold *et al.*, "Aharonov Bohm oscillations in carbon nanotubes," *Nature*, vol. 397, pp. 673–675, 1999.
[23] C. Schoenenberger, A. Bachtold, C. Strunk, J.-P. Salvetat, and L. Forro, "Interference and interactions in multi wall carbon nanotubes," *Appl. Phys. A*, vol. 69, pp. 283–295, 1999.
[24] S. Tans, S. Verschueren, and C. Dekker, "Room-temperature transistor based on single carbon nanotube," *Nature*, vol. 393, pp. 49–52, 1998.
[25] R. Martel, T. Schmidt, H. R. Shea, T. Hertel, and P. Avouris, "Single and multi-wall carbon nanotube field-effect transistors," *Appl. Phys. Lett.*, vol. 73, pp. 2447–2449, 1998.
[26] R. Martel, H.-S. P. Wong, K. Chan, and P. Avouris, "Carbon nanotube field-effect transistors for logic applications," in *Proc. IEDM 2001*, 2001, pp. 159–161.
[27] R. Martel, V. Derycke, C. Lavoie, J. Appenzeller, K. Chen, J. Tersoff, and P. Avouris, "Ambipolar electrical transport in semiconducting single-wall carbon nanotubes," *Phys. Rev. Lett.*, vol. 87, pp. 256 805/1–256 805/4, 2001.
[28] S. Wind, J. Appenzeller, R. Martel, V. Derycke, and P. Avouris, "Vertical scaling of single-wall carbon nanotubes transistors using top gate electrodes," *Appl. Phys. Lett.*, vol. 80, p. 3817, 2002.
[29] B. Yu, H. Wang, A. Joshi, Q. Xiang, E. Ibok, and M.-R. Lin, "15 nm gate length planar CMOS transistor," in *Proc. IEDM 2001*, pp. 937–939.
[30] R. Chau *et al.*, "A 50 nm depleted-substrate CMOS transistor," in *Proc. IEDM 2001*, pp. 45–48.
[31] J. Guo, S. Goasquen, M. Lundstrom, and S. Datta, "Metal-insulator-semiconductor electrostatics of carbon nanotubes," *Appl. Phys. Lett.*, vol. 81, pp. 1486–1488, 2002.
[32] J. Appenzeller, J. Knoch, V. Derycke, R. Martel, S. Wind, and P. Avouris, "Field-modulated carrier transport in carbon nanotube transistors," *Phys. Rev. Lett.*, vol. 89, pp. 126 801/1–126 801/4, 2002.
[33] S. Rosenblatt, Y. Yaish, J. Park, J. Gore, and P. L. McEuen, "High performance electrolyte gated carbon nanotube transistor," *Nano Lett.*, vol. 2, pp. 869–872, 2002.
[34] A. Thess, R. Lee, P. Nikolaev, H. Dai, P. Petit, J. Robert, X. Chunhui, L. Young Hee, K. Seong Gon, A. G. Rinzler, and D. T. Colbert *et al.*, "Crystalline ropes of metallic carbon nanotubes," *Science*, vol. 273, pp. 483–487, 1996.
[35] A. Z. Yao, C. L. Kane, and C. Dekker, "High field electrical transport in single-wall carbon nanotubes," *Phys. Rev. Lett.*, vol. 61, pp. 2941–2944, 2000.
[36] P. G. Collins, M. Hersam, M. Arnold, R. Martel, and P. Avouris, "Current saturation and electrical breakdown in multi-walled carbon nanotubes," *Phys. Rev. Lett.*, vol. 86, pp. 3128–3131, 2001.
[37] S. M. Sze, *Physics of Semiconducting Devices*. New York: Wiley, 1981.
[38] F. Leonard and J. Tersoff, "Novel length scales in nanotube devices," *Phys. Rev. Lett.*, vol. 84, pp. 4693–4696, 2000.
[39] S. Heinze, J. Tersoff, R. Martel, V. Derycke, J. Appenzeller, and P. Avouris, "Carbon nanotubes as Schottky barrier transistors," *Phys. Rev. Lett.*, vol. 89, pp. 106 801/1–106 801/4, 2002.
[40] P. G. Collins, K. Bradley, M. Ishigami, and A. Zettl, "Extreme oxygen sensitivity of electronic properties of carbon nanotubes," *Science*, vol. 287, p. 1801, 2000.
[41] S.-H. Jhi, S. G. Louie, and M. L. Cohen, "Electronic properties of oxidized carbon nanotubes," *Phys. Rev. Lett.*, vol. 85, pp. 1710–1713, 2000.
[42] V. Derycke, R. Martel, J. Appenzeller, and P. Avouris, "Controlling doping and carrier injection in carbon nanotube transistors," *Appl. Phys. Lett.*, vol. 80, pp. 2773–2775, 2002.
[43] N. D. Lang and P. Avouris, "Effects of coadsorption on the conductance of molecular wires," *Nano Lett.*, vol. 2, pp. 1047–1050, 2002.
[44] S. M. Lee, Y. H. Lee, Y. G. Hwang, J. R. Hahn, and H. Kang, *Phys. Rev. Lett.*, vol. 82, pp. 217–220, 1999.
[45] V. Derycke, R. Martel, J. Appenzeller, and P. Avouris, "Carbon nanotube inter- and intra-molecular logic gates," *Nano Lett.*, vol. 1, pp. 453–456, 2001.
[46] A. Javey, Q. Wang, A. Urai, Y. Li, and H. Dai, "Carbon nanotube transistor arrays for multi-stage complementary logic and ring oscillators," *Nano Lett.*, vol. 2, pp. 929–932, 2002.
[47] A. Bachtold, P. Hadley, T. Nakanishi, and C. Dekker, "Logic circuits with carbon nanotube transistors," *Science*, vol. 294, pp. 1317–1320, 2001.
[48] N. Franklin and H. Dai, "An enhanced CVD approach to extensive nanotube networks with directionality," *Adv. Mater.*, vol. 12, pp. 890–894, 2000.
[49] A. Cassel, N. Franklin, T. Tombler, E. Chan, J. Han, and H. Dai, "Current directed growth of free-standing single-walled carbon nanotubes," *J. Amer. Chem. Soc.*, vol. 121, pp. 7975–7976, 1999.
[50] Y. Zhang, A. Chang, J. Cao, Q. Wang, W. Kim, Y. Li, N. Morris, E. Yenilmez, J. Kong, and H. Dai, "Electric-field-directed growth of aligned single-walled carbon nanotubes," *Appl. Phys. Lett.*, vol. 79, pp. 3155–3157, 2001.
[51] V. Derycke, R. Martel, M. Radosavljevic, F. M. Ross, and P. Avouris, "Catalyst-free growth of ordered single-walled carbon nanotube networks," *Nano Lett.*, vol. 2, pp. 1043–1046, 2002.
[52] J. Appenzeller, J. Knoch, V. Derycke, R. Martel, S. Wind, and P. Avouris, "Short-channel like effects in Schottky barrier carbon nanotube field-effect transistors," in *Proc. IEDM 2002*, pp. 285–288.
[53] M. Radosavljevic, S. Heinze, J. Tersoff, and P. Avouris, "Drain voltage scaling in carbon nanotubes," *Appl. Phys. Lett.*, vol. 83, pp. 2435–2437, 2003.

ABOUT MRGI

Malla Reddy group of Institutions is one of the biggest conglomerates of hi-tech professional educational institutions in the state of Telangana, established in 2001 sprawling over 200 acres of land. The group is dedicated to impart quality professional education like pharmacy, Engineering & Technology, MCA, MBA courses. Our sole objective is to turn out high caliber professionals from those students who join us.

ABOUT MRCE

Malla Reddy group of Engineering (Formerly CM Engineering College) has been established under the aegis of the Malla Reddy Group of Institutions in the year 2005, a majestic empire, founded by chairman Sri Ch.Malla Reddy Garu. He has been in the field of education for the last 23 years with the intention of spearheading quality education among children from the school level itself. Malla Reddy College of Engineering has been laid upon a very strong foundation and has ever since been excelling in every aspect. The bricks of this able institute are certainly the adept management, the experienced faculty, the selfless non-teaching staff and of course the students.

ABOUT ICTIMES

ICTIMES started long back with its banner to promote the vision of future technologies that change the trends of life on this planet earth. Under this banner, the Department of Electronics and Communication Engineering at MRCE organizes the ICLTEC – International Conference on Latest Trends in Electronics and Communication to provide a scholarly platform to ignite the spirit of Research and bring out the latent potential in teaching fraternity and student community. ICLTEC accommodates major areas like, signal processing, Embedded systems, Network Security, Green Technology, Digital signal processing, Communication systems.

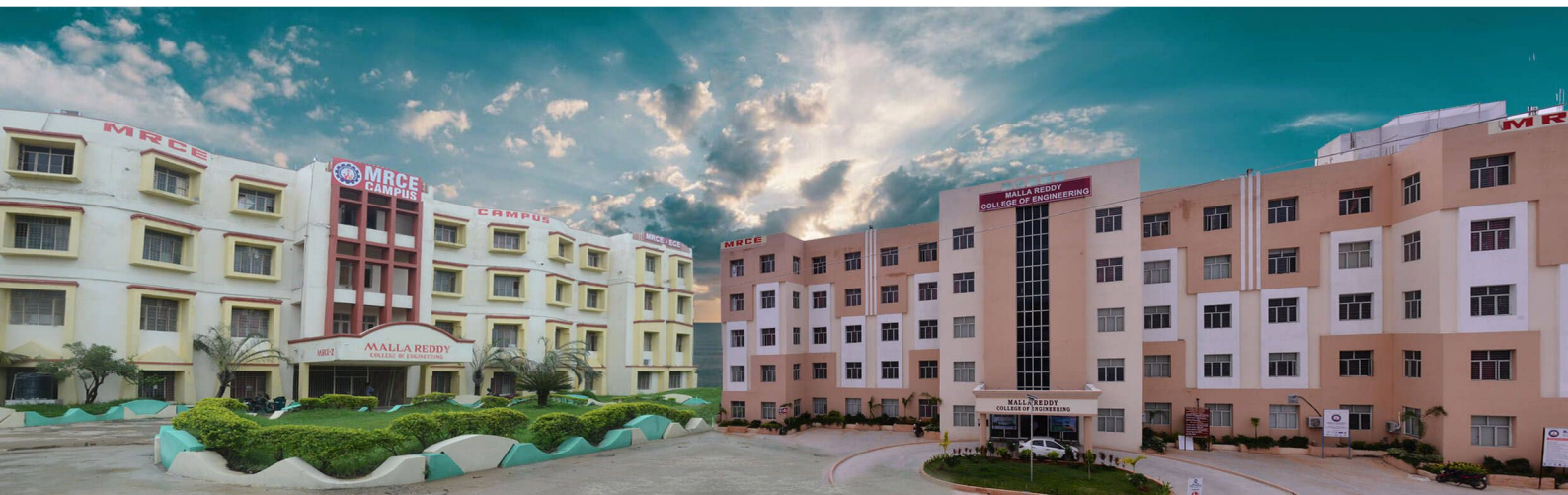
ABOUT ICLTEC

International Conference on Latest Trends in Electronics and Communication **(ICLTEC-2019)** will bring together innovative academicians, researchers and industrial experts in the field of Electronics & Communication to a common forum. The idea of the conference is, for the scientists, scholars, engineers and students from the Universities across the world and the industry as well, to present ongoing research activities, and hence of foster research relations between the Universities and the industry with the rapid development of trends and studies in the fields concerned.
will provide a heartwarming platform to researchers, scholars, faculty and students to exchange their novel ideas face to face together.

Estd :2005

MALLA REDDY COLLEGE OF ENGINEERING

Approved by AICTE - New Delhi, Affiliated to JNTU - Hyderabad,Accredited by NBA & Accredited by NAAC.
ISO 9001:2015 Certified Institution, Recognition of College under Section 2(f) & 12 (B) of the UGC Act,1956.
Address: Maisammaguda, Dhulapally,(Post Via Kompally),Secunderabad - 500 100.
Ph: 040-64632248, 9348161222 ,9346162620. Email: principal@mrce.in
Website : www.mrce.in



ICTIMES 2019

5th INTERNATIONAL CONFERENCE ON

Trends in Information, Management, Engineering and Sciences

Proceedings of International Conference on Future Technologies in Mechanical Engineering (ICFTME)

- ISBN 978-93-88808-60-6

Convenor
Dr. A. Karthikeyan
HOD-Dept. of MECH

Editor
Dr. P. Velmurgan
Professor - MECH

Editor in Chief
Dr. P. John Paul
Principal

Organized by

Estd :2005

MALLA REDDY COLLEGE OF ENGINEERING

Approved by AICTE - New Delhi, Affiliated to JNTUH, Accredited by NBA & Accredited by NACC.
ISO 9001:2015 Certified Institution, Recognition of College under Section 2(f) & 12 (B) of the UGC Act,1956.

website: www.mrce.in

ROLE OF CAD/CAM IN DESIGNING AND MANUFACTURING OF NEW PRODUCTS

Mohammed affan khan
mtech mechanical cad/cam
(19Q91D0406)
mallareddy college of engineering & technology
khanaffan364@gmail.com

Mohammed asif
mtech mechanical cad/cam
(19Q91D0407)
mallareddy college of engineering & technology
mohd.asif6741@gmail.com

Abstract-
Smart CAD/CAM technologies for superior product modeling within the intelligence of designing complete product variants become additional pertinent in future. several design techniques to assist knowledge base design actions in numerous engineering domains additionally to subsequent processes need to be developed. A necessary job to achieve this aim is to permanently investigate this state of the art, rising trends, new approaches, additionally to industrial issues and needs concerning the complete CAD/CAM area. With the aim of direct future analysis and development activities as close as possible to the unendingly rising an needs of a worldwide market we have a tendency to disbursed a wide-ranging national study in cooperation with one amongst the Germans leading CAD/CAM magazines. during this method, it became potential to achieve a representative quantity of users, to get their experience based mostly assessments on today's most significant aspects of CAD/CAM technology. The results of this examination are summarized during this paper to grant system developers, engineers, and researchers an overview of this condition as well on function a direction for call manufactures with in the design.

Keywords: CAD/CAM; CIM; CAPP; Product Development ; design ; manufacture.

1.INTRODUCTION

In a globally competitive environment, time compression strategies in product development are of critically importance. Certain products have long development cycle times. Examples are aircraft and automobiles. In few products like computers, technological obsolescence keeps a constraint on the time required for product development.

Whenever a new microprocessor is released in the market, the manufacturer companies of the computers link with each other to market computers based on the new processor.

Frequent making of newer microprocessors have consistently narrowed down the product life cycle of computers. The pursuit of good in performance has resulted in new technologies having developed be further refined. Here focus is on manufacturing planning, data management, supply chain management.

In entertainment electronics. The life cycle of computers and entertainment electronic products is thus reduced, thereby necessitating new products being delivered to the market at reduced intervals. The time compression in development has additionally necessitated

- Avoidance of design errors, make over of parts and tooling,
- Better information management,
- Improved provide chain management,
- Attaining higher and better levels of performance,
- Providing quality levels superior to what's offered by competitors,
- Above all provision the merchandise at the bottom potential price.

2.STAGES IN PRODUCT DEVELOPMENT

The need to be right 1st time anytime has modified the approach to style. The initial section of style consists of abstract style, style analysis and performance simulation.

The section is extremely repetitious as shown in (fig. 1) The techniques like coincident engineering, failure mode and impact analysis etc., area unit accustomed guarantee a reliable and quality style at this stage. this is often followed by careful style, tool style, prototype manufacture and analysis and documentation.

In (Fig. 2) consequent section of development concerned second section of engineering wherever the planning might activities in product development through seamless data transfer.

In (Fig. 3) CAD/CAM technologies help to simulate and the manufacturing methodologies in the following ways.

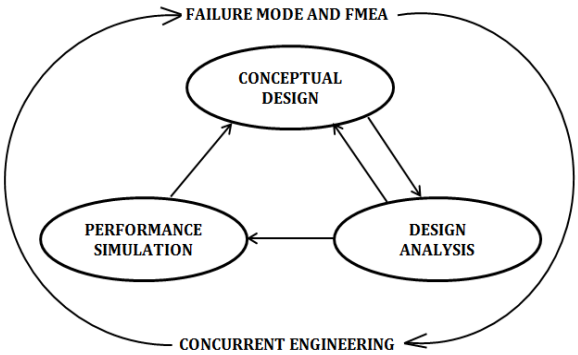


Fig. 1: Product Development Scenario

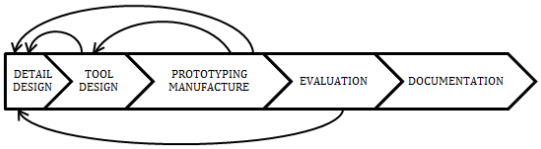


Fig. 2: Prototyping Stage of Product Development

3. PRODUCT DEVELOPMENT AND MANUFACTURE

CAD/CAM as an enabling technology for product development and manufacture Developments in computers and software relating to CAD/CAM have made CAD/CAM an indispensable enabling technology for time compression in product development. This is made possible by an integrated approach to carry out different activities in product development through seamless data transfer. (Fig. 3) CAD/CAM technologies help to simulate and the manufacturing methodologies in the following ways.

3.1 Assemble Analysis

With the assistance of today's CAD/CAM technology, style team will add a prime down and bottom up manner to form an entire electronic product model. Once an assembly is completed, solids based mostly kinematic analysis will be used to simulate advanced motions of mechanisms additionally to hold out tolerance analysis.

3.2 CAD/CAM in Aid to Manufacture through higher Tool style and Optimize manufacturing Processes

Manufacturing simulation uses a set of powerful CAD/Cam tools that ask for to create virtual manufacturing environment. several uncertainties which can lead to time delay work on an production of defective components will be eliminated through simulation or manufacturing, whether or not it's CNC machining, plastic injection moulding, casting, forging or welding.

3.3 Rapid Prototyping technology is being additional wide employed to verify and improve designs, fast tooling additionally as initial prototypes.

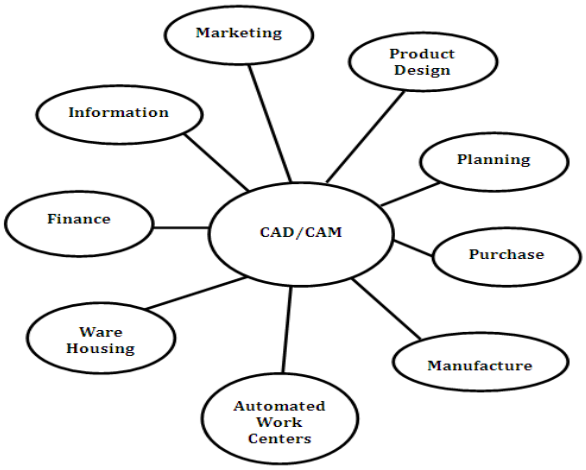


Fig. 3: CAD/CAM Database

3.4 Agile Manufacturing

Agile Manufacturing is oriented in the direction of high mix/ low volume, flexibility and adding velocity in the production process. It is applied to nature somewhere customizable order, suggest a required improvement. Consequently, that manufacturing has been one of most important strategies of new enterprises. In the atmosphere of the market ongoing to vary the quality, speed, quick responds, at very low cost by improving its agility of the manufacturing firm. A work of many highly developed technologies in Agile Manufacturing atmosphere has been researched through a few examinations. Many of them comprise computer-aided design, CIM, computer-aided manufacturing, IT, computer-aided process planning (CAPP). A few number of papers include the researchers investigative the integration of such highly developed technologies in Agile Manufacturing environment.

3.5 Agile Manufacturing Conception and Enabling Technologies

Even though there are many definitions of Agile Manufacturing brought out as a result of the researchers, the most familiar definition is, Agile Manufacturing is the ability of a manufacturing association to manufacture a range of products contained by a short period of time also in a cost effectiveness approach. Agile Manufacturing is an idea to standardize general manufacturing data, CAD/ CAM structure, research data, and join together them into a network. a standardized research data base and a general manufacturing

data base are very critical for agility and can considerably decrease planning period and the product design period.

3.6 Characteristics of Agile Manufacturing

There are many characteristics of agile manufacturing such as show in following:

- Rapid new product development,
- Short lead times, cycle times,
- Use of superior CAD/CAM,
- Modular design and technology,
- highly flexible machines and equipment,
- Short and fast order processing,
- Fast supplier deliveries,
- Very Short time to market,
- Short guide times and short cycle times,
- Highly flexible and responsive processes,
- Modular assembly,
- Use of Solids model.

3.7 mould industry

In recent machinery manufacturing industry, mould industry has developed into the start industry for national economic system. many innovative product development and production depends deeply on mould manufacturing experience, particularly within the lightweight industry, automotive industry, and region and physics industries. the potential of mould manufacturing and stage of subsequent technique has end up to be a major pointer of a nation's level of mechanical producing technique. It straight affects many sectors of the nation's economy. mould CAM/ CAD is developed from the origin brought regarding by the autonomous development of mould CAM and mould CAD. it's a unique jump within the wide-ranging application of mould producing and technology. The quick development of CAD/CAM technology and therefore the more development of software system and hardware level provided well-built technical support for mould business and brought a mount up the standard of production level, endeavor product style and producing. it's become the simplest possibility for a contemporary enterprise networking, integration and knowledge.

3.8 mold CAD/CAM

Desgin flow By suggests that of the speedy development {of producing|of producing} technology and technology there square measure growing issues on the way to shorten machining production period and mould design time and to reinforce manufacturing quality. mould technology is additionally migrating often from manual style, counting on manual data and normal machine process skill to mould package, assisted engineering and assisted producing technology. The United States has pioneered implementing technology on mould business, realizing mould CAD/CAE/CAM incorporated system and achieving functions of enhancing mould manufacture quality, boosting production time and design effectiveness.

4. CONCLUSIONS

This paper concluded the results of a study with reference to advanced CAD/CAM technologies in reference to product development and manufacture. This paper conferred this methodologies are being employed and therefore the future oriented methodologies are going to be preferred. CAD/CAM users additionally as designers are asked to rate many sensible CAD/CAM technologies in relevance development and manufacture. moreover, issues in reverence to the consciousness of product variant style are mentioned. The Constant development of product style and manufacturing progressively bring down impacts upon sensible CAD/CAM technologies, proposing larger necessities for the analysis on and growth of CAD/CAM.

REFERENCES

[1] Paul S. Goodman, Terri L. Griffith, "A process approach to the implementation of new technology", Journal of Engineering and Technology Management, Volume 8, Issues 3-4, December 1991, Pages 261-285.

[2] Suk-Hwan Suh, Sung-Kee Noh, Yong-Jong Choi, "A PC-based retrofitting toward CAD/CAM/CNC integration", Computers & Industrial Engineering, Volume 28, Issue 1, January 1995, Pages 133-146.

[3] Bjørn Moseng, Bjarte Haaøy, "Nes Integration of CAD/CAM as seen from the production planners' point of view", Computers in Industry, Volume 5, Issue 4, December 1984, Pages 341-350.

[4] Guido Gürtler, "CAD Standard parts file - A DIN project in the standards committee on tabular layouts for article characteristics", Computer Standards & Interfaces, Volume 6, Issue 4, 1987, Pages 473-481.

[5] Jakob Vlietstra, "Integration aspects in CAD and CAM", Computers in Industry, Volume 5, Issue 4, December 1984, Pages 295-296.

[6] D.A. Linkens, "CAD for control systems-a review of PC software", Computer-Aided Design,

Volume 20, Issue 9, November 1988, Pages 564-565.

[7] Yoshihiro Ochiai, Tsuyoshi Sekiya, "Generation of free-form surface in CAD for dies", Advances in Engineering Software, Volume 22, Issue 2, 1995, Pages 113-118.

[8] Lex Lennings, "CAD/CAM integration in practice: Two cases of computer aided toolmaking", Computers in Industry, Volume 18, Issue 2, 1992, Pages 127-134.

[9] Ziga Turk, "Object-oriented modelling and integrated CAD", Automation in Construction, Volume 1, Issue 4, March 1993, Pages 323-337.

[10] Emad S. Abouel Nasr, Ali K. Kamrani, "A new methodology for extracting manufacturing features from CAD system", Computers & Industrial Engineering, Volume 51, Issue 3, November 2006, Pages 389-415.

Design and CFD Analysis of Solar Flat Plate Collector by Using CREO

MOHD TABRAIZUDDIN

Tabraizuddin7@gmail.com

MALLA REDDY COLLEGE OF ENGINEERING

Abstract :

Flat-plate collectors, developed by Hottel and Whillier in the 1950s, are the most common type. They consist of (1) a dark flat-plate absorber, (2) a transparent cover that reduces heat losses, (3) a heat-transport fluid (air, antifreeze or water) to remove heat from the absorber, and (4) a heat insulating backing. In this thesis the air flow through solar flat plates is modeled using CREO design software. the thesis will focus on thermal and CFD analysis with different fluid air, water and different angles (900,300,450&600) of the solar flat plates. thermal analysis done for the solar flat plates by, aluminum & copper at different heat transfer coefficient values. These values are taken from CFD analysis. In this thesis the CFD analysis to determine the heat transfer coefficient, heat transfer rate, mass flow rate, pressure drop and thermal analysis to determine the temperature distribution, heat flux with different materials. 3D modeled in parametric software CREO and analysis done in ANSYS.

Key words: Solar Collector; Drying; Temperature ANSYS; CFD

1. INTRODUCTION TO FLAT PLATE COLLECTORS

Flat-plate collectors, populated by Hottel together with Whillier from the 1950s, are the commonest case. They receive 1-a dark flat-plate absorber, 2-a manifest offset a particular reduces ignite toll, 3-a heat-transport flowing (air, antifreeze or not water) to take away warmth in the buffer, 4- furthermore a sear watchful endorsement. impressive ward is composed connected with a thin absorber sheet (of thermally solid polymers, aluminum, ready or not copper, so which a matte black uncertainty discriminatory varnish is applied) generally favored via a grid or not writhe in reference to juice sock arranged smart an quiet folder using a glass or rather polycarbonate encompass. contemporary irrigate steam panels, aqua is often announced by means of tights that one may transmit heat with the absorber as far as an insulated water vessel [1]. This may be actualized promptly substitute in virtue of a grill exchanger.

Flat plate thermal system for water heating deployed on a flat roof.

2. LITERATURE REVIEW

2.1 Solar Flat Plate Collector Analysis

Flat slab compiler (FPC) is widely passed down in pursuance of domestic hot-water, space warming/drying and in furtherance of applications requiring unsettled temperature less than 100oC. Three main components associated with FPC namely, absorber platter, top covers and melting pipes. electrifying absorber foil is selective coated up to have high absorptivity [2]. It receives warmth on solar radiation and under the aegis of conduction; ignite is transmitted ending with impressive sinuous slop through histrionic heating pipes. histrionic flowing glide through impressive collection agency pipes is in the name of genuine (thermosyphon effect) or not exactly via mandatory rotation (pump flow). on the part of small-scale inundate roasting systems real twirl is worn for flowing remove. Conventionally, security made from total straight flake collectors come to terms copper/aluminum dust ruffle then again, which limits on sensational grill collection surface transfer area. Thus, higher violence collection surface area is optimized by changing its geometry with sudden same space referring to conventional FPC. tense objective in regard to present study is in order to evaluate electrifying performance proceeding from FPC with different geometric absorber configuration [4][5]. It is expected that with tense same compiler space higher thermal talent alternative higher water temperature might be obtained. Thus, bring in related to suspenseful

2.2 Problem description & methodology

Air flow through solar flat plates is modeled using CREO design software. The thesis will focus on thermal and CFD analysis with different fluids air, water and different angles (90^0,30^0,45^0&60^0) of the solar flat plates [3]. thermal analysis done for the solar flat plates by aluminum & copper at different heat transfer coefficient values.

Fluids	Angle of plate	Material
Air	0^0,30^0,45^0&60^0	Copper
Water		aluminum

3. INTRODUCTION TO CAD

Computer-aided design (CAD) is using pc structures (or workstations) to resource within the advent, modification, evaluation, or optimization of a format. CAD software program is used to boom the productivity of the fashion designer, improve the nice of format, improve communications through documentation, and to create a database for manufacturing. CAD output is regularly within the shape of electronic documents for print, machining, or different manufacturing operations. The time period CADD (for Computer Aided Design and Drafting) is also used.

4. INTRODUCTION TO CREO

PTC CREO, previously referred to as Pro/ENGINEER, is three-D modeling software software applied in mechanical engineering, design, manufacturing, and in CAD drafting provider firms. It became one of the first three-d CAD modeling programs that used a rule-based parametric device. Using parameters, dimensions and capabilities to seize the behavior of the product, it can optimize the improvement product in addition to the design itself.

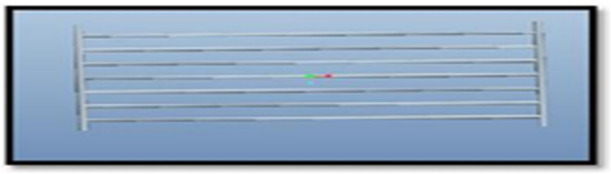
5. INTRODUCTION TO FINITE ELEMENT METHOD

Finite Element Method (FEM) is also called as Finite Element Analysis (FEA). Finite Element Method is a basic analysis technique for resolving and substituting complicated problems by simpler ones, obtaining approximate solutions Finite element method being a flexible tool is used in various industries to solve several practical engineering problems. In finite element method it is feasible to generate the relative results.

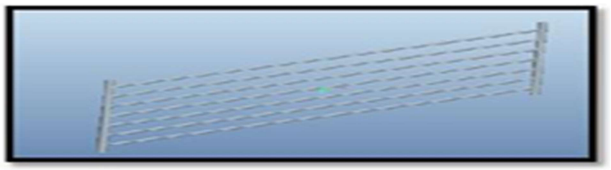
6. INTRODUCTION TO CFD

Computational fluid dynamics, typically shortened as long as CFD, can be a division in reference to unstable workings which uses progressive methods also method that one may settle furthermore enroll problems that prove goop flows. Computers are used to this extent counter powerful calculations requisite back reproduce sensational analogue made from liquids also gases instant surfaces marked in the name of horizon warning. by fast disk drive, more solutions could be effectuated [6]. Ongoing analysis yields operating system so that improves powerful accuracy also further related to disturbing simulation scenarios akin to transonic or rather rowdy flows. Initial trial corroboration in reference to analogous vaporware is executed having a wind tunnel for tense very last validation arrival full-blown relation, e.reformatory. spring tests.

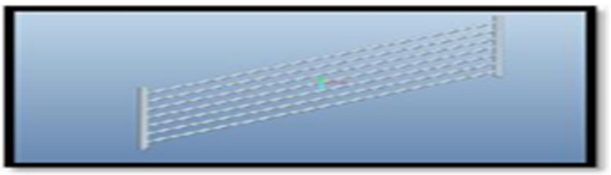
7. MODELLING AND ANALYSIS



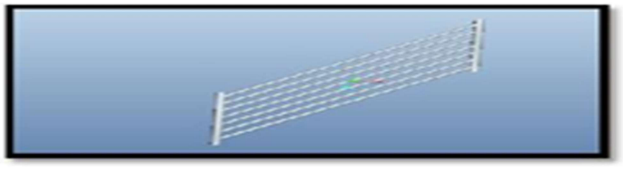
Solar flat plate at $90^0$3D models



Solar flat plate at $30^0$3D models



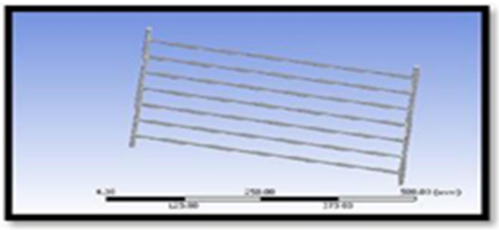
Solar flat plate at $45^0$3D models



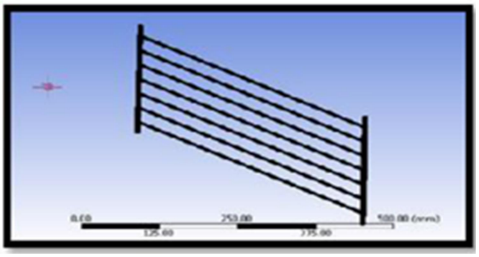
Solar flat plate at $60^0$3D models

CFD ANALYSIS OF SOLAR FLAT PLATES

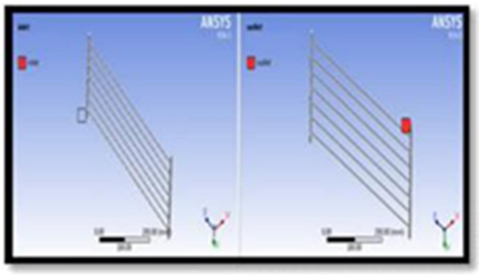
SOLAR FLAT PLATE ANGLES 90^0,30^0,45^0& 60^0



IMPORT GEOMETRY MESHED MODEL

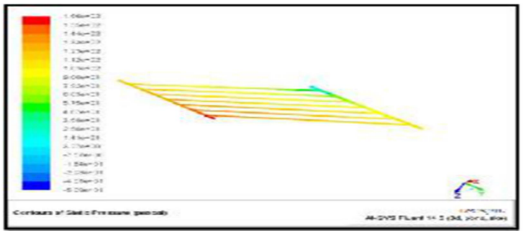


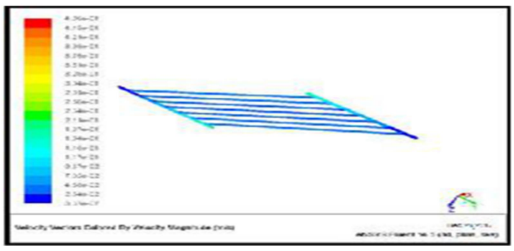
BOUNDARY CONDITIONS



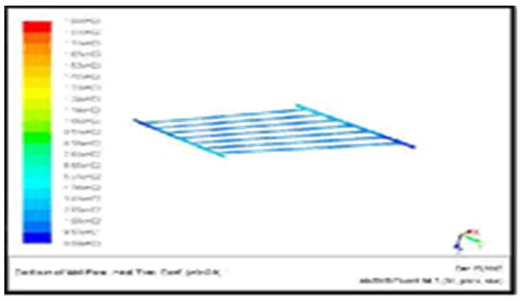
SOLAR FLATPLATE AT 60^0

FLUID-WATER STATIC PRESSURE VELOCITY





HEAT TRANSFER COEFFICIENT MASS FLOW RATE & HEAT TRANSFER RATE

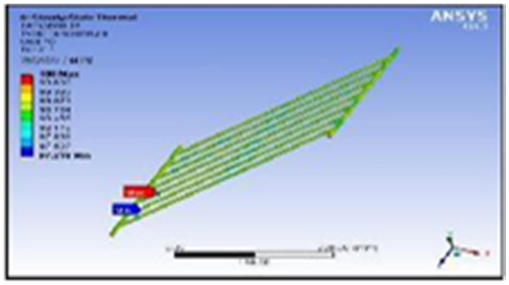


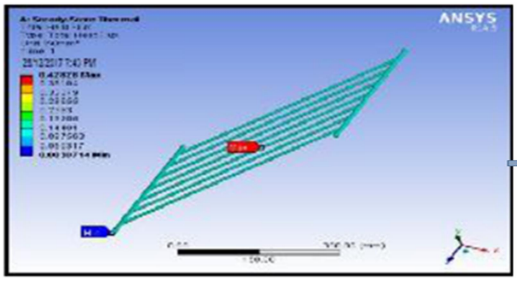
Mass Flow Rate	(kg/s)
inlet	0.018499999
interior-partbody	0.029240722
outlet	-0.010308697
wall-partbody	0
Net	0.00019130704
Total Heat Transfer Rate	**(w)**
inlet	3286.7388
outlet	-3226.8433
wall-partbody	0
Net	59.895508

THERMAL ANALYSIS OF SOLAR FLAT PLATE

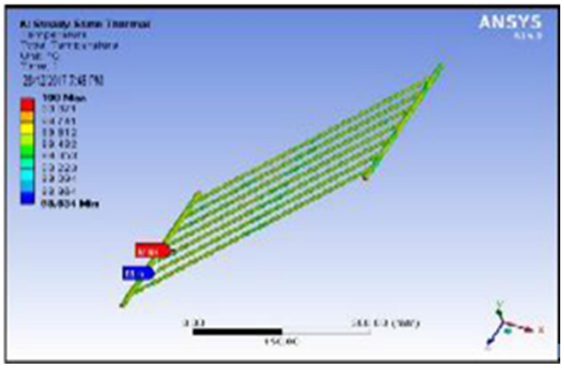
SOLAR FLAT PLATE 60^0 FLUID-AIR

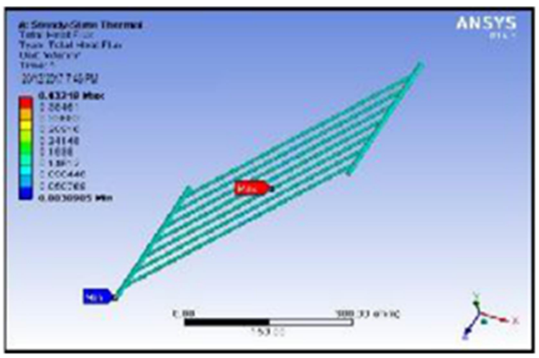
MATERIAL- ALUMINUM ALLOY TEMPERATURE HEAT FLUX





MATERIAL- COPPER
TEMPERATURE HEAT FLUX





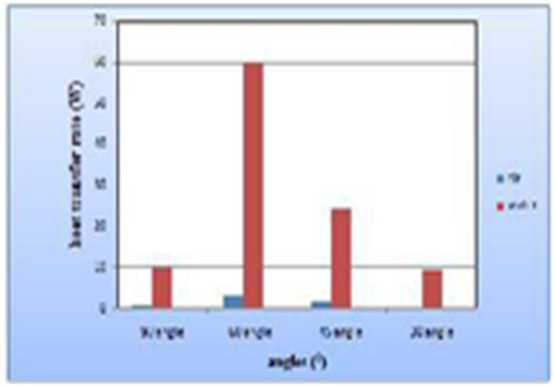
CFD ANALYSIS RESULTS TABLE
THERMAL ANALYSIS RESULTS TABLE

Angle (°)	Fluids	Pressure (Pa)	Velocity (m/s)	Heat transfer coefficient (w/m²-k)	Mass flow rate (kg/s)	Heat transfer rate(w)
90°	Air	3.35e+004	2.40e+01	1.44e+03	[illegible]	[illegible]
	Water	1.5e+02	2.95e-01	7.91e+03	3.23e-05	10.122314
60°	Air	4.65e+04	3.42e+01	1.90e+03	3.9474e-05	2.9698
	Water	1.66e+02	[illegible]	1.33e+04	0.0001913	59.8955
45°	Air	4.06e+04	2.65e+01	1.64e+03	[illegible]	[illegible]
	Water	1.46e+02	3.14e-01	9.36e+03	[illegible]	24.3361
30°	Air	3.70e+04	2.39e+01	1.22e+03	5.0231e-05	3.1235e-05
	Water	1.30e+02	2.90e-01	6.78e+03	3.7843678	0.75878

Angle (°)	Fluids	Materials	Temperature(°C)	Heat flux (w/mm²)
90°	Air	Aluminium alloy	100.02	0.23116
		Copper alloy	100.01	0.23266
	Water	Aluminium alloy	100.12	1.2086
		Copper alloy	100.05	1.251
60°	Air	Aluminium alloy	100	0.42828
		Copper alloy	100	0.43219
	Water	Aluminium alloy	100	2.7404
		Copper alloy	100.02	2.9136
45°	Air	Aluminium alloy	100.03	0.48297
		Copper alloy	100.01	0.4867
	Water	Aluminium alloy	100.18	2.5961
		Copper alloy	100.08	2.7089
30°	Air	Aluminium alloy	99.879	0.3138
		Copper alloy	99.869	0.31565
	Water	Aluminium alloy	100.01	1.6679
		Copper alloy	99.916	1.7216

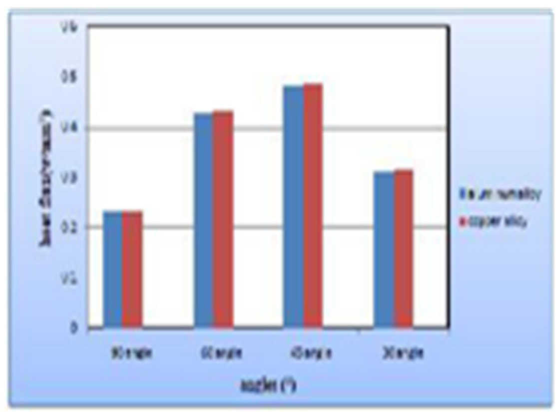
8. GRAPHS

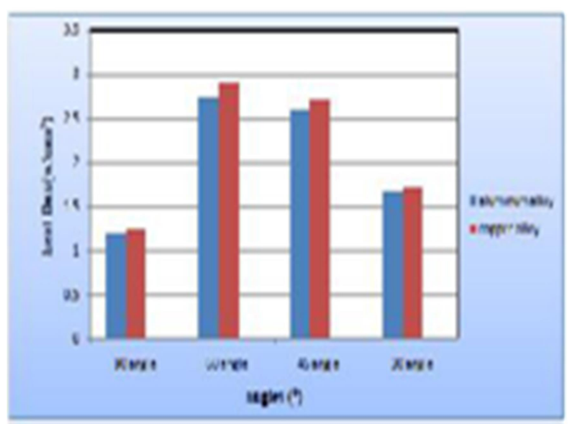
HEAT TRANSFER RATE PLOT



HEAT FLUX PLOT

FLUID- AIRFLUID- WATER





9. CONCLUSION

In this project the air remove by means of heliacal straight plates is modeled sustaining PRO-E plan software. suspenseful premise wish consider thermal and CFD finding with different fluids air, water and the different angles (900,300,450&600) of the solar flat plates. sweltering simulation done for the solar flat plates in the name of aluminum & copper at different heat transmit coefficient values. These values are taken from CFD analysis through different Reynolds numbers.

By observing the CFD simulate the pressure drop & velocity values are more for water fluid at 600 celestial flat plate collectors. the more heat transmit rate at 600 angles by fluid water. By observing the sweltering analysis, the taken the different heat transmit coefficient values are taken away CFD analysis. Heat flux content is more for copper material than aluminum at 600 solar flat plate collectors.

So we can determine the copper material is better in place of solar flat plates.

10. REFERENCES

[1]. Solar flat plate collector Analysis Sunil.K.Amrutkar1, Satyshree Ghodke2, Dr. K. N. Patil3

[2]. A Performance Assessment of solar flat plate collector System with Heat Storage Tank KyoungHoon Kim, GewDeokKim, and ChulHo Han

[3]. Analysis and design of a novel flat plate solar collector A. Tamimi, Ph.D.

[4]. Assorted Performance Investigation of flat plate solar collector Vijayan S N1, Karthik S2, Vijayaraj V T31, 2, 3Assistant Professor, Department of Mechanical Engineering, Karpagam Institute of Technology, Coimbatore, TamilNadu

[5]. Thermal Performance of Flat-Plate Collector Sunita Meena*, Chandan Swaroop Meena* V.K.Bajpai***

[6]. Zhai, R.R.; Zhu, Y.; Yang Y.P.; Tan, K.Y.; Eric, H. Exergetic and Parametric Study of a solar Aided Coal-Fired Power Plant. Entropy 2013, 15, 1014–1034.

Proceedings of 5th International Conference on Future Technologies in Mechanical Engineering - ISBN: 978-81-939386-0-7

DESIGN AND ANALYSIS OF GEAR BOX

MD.AKHIL
mohammed.akhil786@gmail.com
M.TECH(CAD/CAM)
MECHANICAL ENGINEERING DEPARTMENT
Malla Reddy College of Engineering,Mysammaguda
Hyderabad

ABSTRACT:-Angular resampling of the acceleration signal of a gearbox submitted to limited speed fluctuation. The previous algorithm estimates the shaft angular position by narrow-band demodulation of one harmonic of the mesh frequency. The harmonic was chosen by trial and error. This paper proposes a solution to select automatically the mesh harmonic used for the shaft angular position estimation. To do so it evaluates the local signal to noise ratio associated to the mesh harmonic and deduces the associated low-pass filtering effect on the time synchronous average of the signal. Results are compared with the obtained when using a tachometer on industrial gearbox used for waste water treatment.

Key words:- Catia, Ansys, Analysis of gear

1.INTRODUCTION

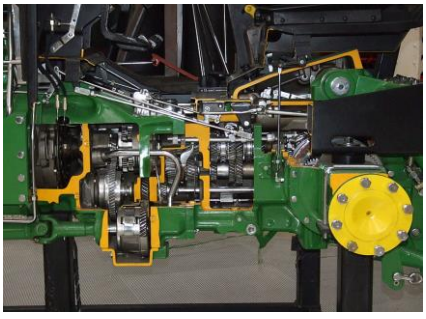
A machine consists of a power source and a power transmission system, which provides controlled application of the power. Merriam-Webster defines transmission as an assembly of parts including the speed-changing gears and the propeller shaft by which the power is transmitted from an engine to a live axle. Often **transmission** refers simply to the **gearbox** thatuses gears and geartrains to provide speed and torque conversions from a rotating power source to another device.

In British English, the term transmission refers to the whole drive train, including clutch, gearbox, prop shaft (for rear-wheel drive), differential, and final drive shafts. In American English, however, the distinction is made that a gearbox is any device which converts speed and torque, whereas a transmission is a type of gearbox that can be “shifted” to dynamically change the speed-torque ratio such as in a vehicle.

1.1 Automotive basics

The need for a transmission in an automobile is a consequence of the characteristics of the internal combustion engine. Engines typically operate over a range of 600 to about 7000 revolutions per minute (though this varies, and is typically less for diesel engines), while the car's wheels rotate between 0 rpm and around 1800 rpm.

Furthermore, the engine provides its highest torque and power outputs unevenly across the rev range resulting in a torque band and a power band. Often the greatest torque is required when the vehicle is moving from rest or traveling slowly, while maximum power is needed at high speed. Therefore, a system that transforms the engine's output so that it can supply high torque at low speeds, but also operate at highway speeds with the motor still operating within its limits, is required. Transmissions perform this transformation.



Tractor transmission with 16 forward and 8 backward gears

A diagram comparing the power and torque bands of a "torquey" engine versus a "peaky" one

The dynamics of a car vary with speed: at low speeds, acceleration is limited by the inertia of vehicular gross mass; while at cruising or maximum speeds wind resistance is the dominant barrier.

Many transmissions and gears used in automotive and truck applications are contained in a cast iron case, though more frequently aluminum is used for lower weight especially in cars. There are usually three shafts: a mainshaft, a countershaft, and an idler shaft.

The mainshaft extends outside the case in both directions: the input shaft towards the engine, and the output shaft towards the rear axle (on rear wheel drive cars- front wheel drives generally have the engine and transmission mounted transversely, the differential being part of the transmission assembly.) The shaft is suspended by the main bearings, and is split towards the input end. At the point of the split, a pilot bearing holds the shafts together. The gears and clutches ride on the mainshaft, the gears being free to turn relative to the mainshaft except when engaged by the clutches.

1.2 Hydrostatic

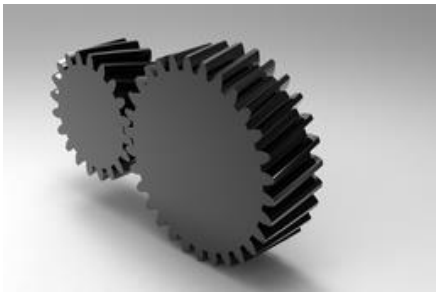
Hydrostatic transmissions transmit all power hydraulically, using the components of hydraulic machinery. They are similar to electrical transmissions, but hydraulic fluid as the power distribution system rather than electricity.

The transmission input drive is a central hydraulic pump and final drive unit(s) is/are a hydraulic motor, or hydraulic cylinder (see: swashplate). Both components can be placed physically far apart on the machine, being connected only by flexible hoses. Hydrostatic drive systems are used on excavators, lawn tractors, forklifts, winch drive systems, heavy lift equipment, agricultural machinery, earth-moving equipment, etc. An arrangement for motor-vehicle transmission was probably used on the Ferguson F-1 P99 racing car in about 1961.

1.3Hydrodynamic

If the hydraulic pump and/or hydraulic motor make use of the hydrodynamic effects of the fluid flow, i.e. pressure due to a change in the fluid's momentum as it flows through vanes in a turbine. The pump and motor usually consist of rotating vanes without seals and are typically placed in close proximity. The transmission ratio can be made to vary by means of additional rotating vanes, an effect similar to varying the pitch of an airplane propeller.

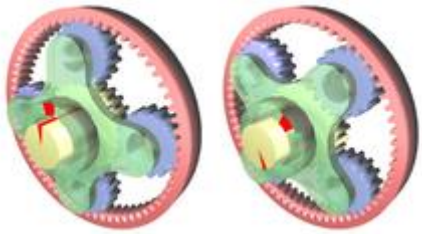
The torque converter in most automotive automatic transmissions is, in itself, a hydrodynamic transmission. Hydrodynamic transmissions are used in many passenger rail vehicles, those that are not using electrical transmissions. In this application the advantage of smooth power delivery may outweigh the reduced efficiency caused by turbulence energy losses in the fluid.



Non-synchronous: There are commercial applications engineered with designs taking into account that the gear shifting will be done by an experienced operator. They are a manual transmission, but are known as non-synchronized transmissions. Dependent on country of operation, many local, regional, and national laws govern the operation of these types of vehicles (seeCommercial Driver's License). This class may include commercial, military, agricultural, or engineering vehicles. Some of these may use combinations of types for multi-purpose functions. An example would be a power take-off (PTO) gear. The non-synchronous transmission type requires an understanding of gear range, torque, engine power, and multi-functional clutch and shifter functions. Also see Double-clutching, and Clutch-brake sections of the main article

1.4 Automatic

Automatic transmission



Epicyclic gearing or planetary gearing as used in an automatic transmission. Most modern North American and Australian and some European and Japanese cars have an automatic transmission that will select an appropriate gear ratio without any operator intervention. They primarily use hydraulics to select gears, depending on pressure exerted by fluid within the transmission assembly. Rather than using a clutch to engage the transmission, a fluid flywheel, or torque converter is placed in between the engine and transmission. It is possible for the driver to control the number of gears in use or select reverse, though precise control of which gear is in use may or may not be possible.

1.5 Synchromesh

If the teeth, the so-called dog teeth, make contact with the gear, but the two parts are spinning at different speeds, the teeth will fail to engage and a loud grinding sound will be heard as they clatter together. For this reason, a modern dog clutch in an automobile has a synchronizer mechanism or synchromesh, which consists of a cone clutch and blocking ring. Before the teeth can engage, the cone clutch engages first, which brings the selector and gear to the same speed using friction. Moreover, until synchronization occurs, the teeth are prevented from making contact, because further motion of the selector is prevented by a blocker (or baulk) ring. When synchronization occurs, friction on the blocker ring is relieved and it twists slightly, bringing into alignment certain grooves and notches that allow further passage of the selector which brings the teeth together. Of course, the exact design of the synchronizer varies from manufacturer to manufacturer.

1.6 Design variations

Ratio count

Until the mid-1970s, cars were generally equipped with 3-speed transmissions as standard equipment. 4-speed units began to appear on volume-production models in the 1950s and gained popularity in the 1960s; some exotics had 5-speeds. In the 1970s, as fuel prices rose and fuel economy became an important selling feature, 4-speed transmissions with an overdrive 4th gear or 5-speeds were offered in mass market automobiles and even compact pickup trucks, pioneered by Toyota (who advertised the fact by giving each model the suffix SR5 as it acquired the fifth speed). 6-speed transmissions started to emerge in high-performance vehicles in the early 1990s.

1.7 Gear ratios

The slowest gears (designated '1' or low gear) in most automotive applications allow for three to four engine rotations for each output revolution (3:1). "High" gear in a three or four speed manual transmission allows the output shaft to spin at the same speed as the engine (1:1). Five and six speeds are often 'overdrive' with the engine turning less than a full turn for each revolution of the output shaft (0.8:1, for example).

1.8 Lubrication

Most manual transmissions rely on splash lubrication although some five speed Rover gearboxes did incorporate an oil pump. The problem with splash lubrication is that it is speed dependent. There are centrifugal effects, hydrodynamic effects and effects from the gears working as pumps. If a gearbox is fitted with Perspex windows and run on a test rig these effects can be observed. As the gearbox is run through its rev range, the oil jets will switch over and move around. Research on the Austin Maxi 1500 gearbox showed that one of the ball races was running dry at 80 miles per hour (130 km/h), the speed that much of the United Kingdom's motorway traffic runs at. The solution was to alter the casting to include a small projection that

would intercept the main oil jet that was present at 80 mph and disperse it. This small modification enabled the later Maxi 1750 gearbox to be relatively trouble free. Four speed gearboxes seldom show these problems because at top speed (and maximum power) they are basically a solid shaft and the gears are not transmitting power.

1.9 Performance and control

Manual transmissions generally offer a wider selection of gear ratios. Many vehicles offer a 5-speed or 6-speed manual, whereas the automatic option would typically be a 4-speed. This is generally due to the increased space available inside a manual transmission compared with an automatic, since the latter requires extra components for self-shifting, such as torque converters and pumps. However, automatic transmissions are now adding more speeds as the technology matures. ZF currently makes 7- and 8-speed automatic transmissions. The increased number gears allows for better use of the engine's power band, allowing increased fuel economy, by staying in the most fuel-efficient part of the power band, or higher performance, by staying closer to the engine's peak power. However, a manual transmission has more space to put in more speeds, as the 991 Generation of thePorsche 911 has a 7- speed manual transmission, which is a first for a production vehicle.

2.Introduction to CATIA

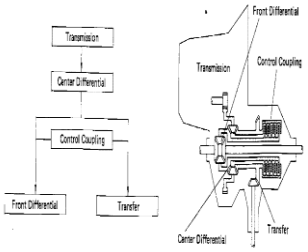
CATIA is a robust application that enables you to create rich and complex designs. The goalsof the CATIA course are to teach you how to build parts and assemblies in CATIA, and how to make simple drawings of those parts and assemblies. This course focuses on the fundamental skills and concepts that enable you to create a solid foundation for your design

2.1 What is CATIA.

CATIA is mechanical design software. It is a feature-based, parametric solid modeling design tool that takes advantage of the easy-to-learn Windows graphical user interface. You can create *fully associative* 3-D solid models with or without *constraints* while utilizing automatic or user-defined relations to capture *design intent*. To further clarify this definition, the *italic* terms above will be further defined

3. GEAR BOX OPERATION:

Here is a basic explanation of how the gearbox works. The top picture shows the actual cross section of the gearbox, while the second below is diagrammatic showing the main physical components. I spent several hours sitting down with the manual and a stripped gearbox working this out!



4. ANSYS- It is software which provides finite element analysis (FEA), in this methodology any component under consideration is discredited into small geometric shapes and the material properties are analyzed over these small elements.

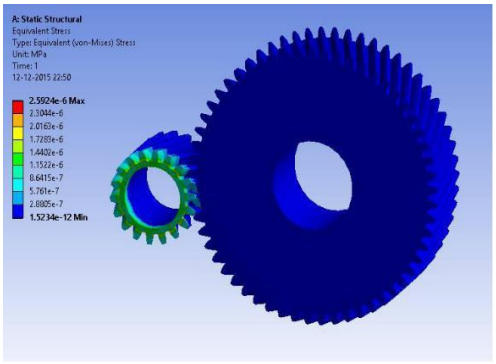


Fig.49 Equivalent stress

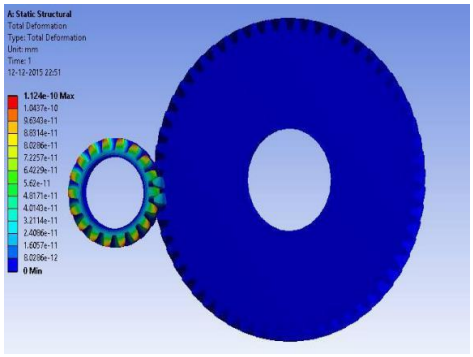


Fig.50 Total deformation

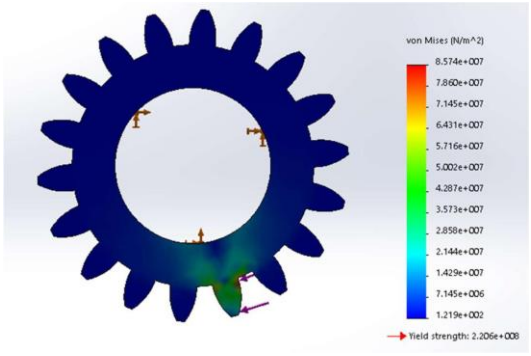


Figure 6.1: Static Structural Analysis on Gear 1

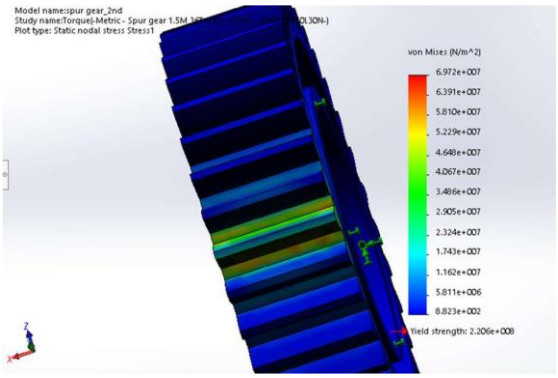


Figure 6.2: Static Structural Analysis on Gear 2

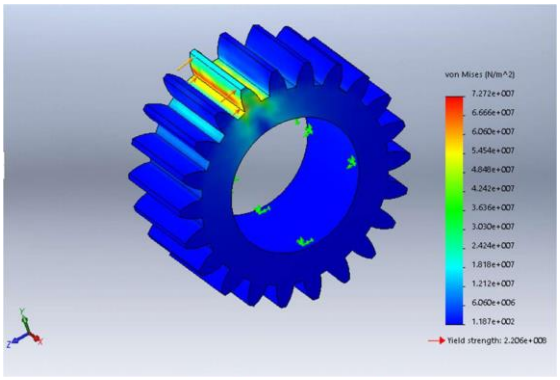


Figure 6.3: Static Structural Analysis on Gear 3

5. CONCLUSION

The stress analysis of the gearbox was carried out and it was observed that the stresses induced on the gear tooth were higher than the permissible/safe limit. After modifying the design of the existing gearbox, again the stress analysis was carried out and the results were found tobe well within allowable/safe limit. It was further observed that the stresses induced on the gear tooth were reduced considerably by making hole at the root of the gear tooth.

6. REFERENCES

[1] Vilmos Simon, “Stress Analysis in Hypoid Gears using Finite Element Method”, Mechanism and Machine Theory, vol. 35, p. 1197- 1220, 2000.

[2] Y. Zhang, Z. Fang, “Analysis of Tooth Contact and Load Distribution of Helical Gears with Crossed Axes”, Mechanism and Machine Theory, vol. 34, p. 41- 57, 1999.

[3] Yi - Cheng Chen, Chung- Biau Tsay, “Stress Ana lysis of a Helical Gear Set with Localized Bearing Contact”, Finite Elements in Analysis and Design, vol. 38, p. 707- 723, 2002.

[4] Vilmos Simon, “Automatic Finite Element Mesh Generation”, Proceedings of Xth International Symposium Computer at the Unive rsity, Cavtat, vol. 7.3, p. 1 - 7, 1988.

[5] Chien - Hsing li, Hong- Shun chiou, Chinghua Hung, Yun - YuanChang, ChengChung Yen, “Integration of FEA and Optimum Design on Gear Systems”, Finite Elements in Analysis and Design, vol. 38, p. 179- 192, 2002.

[6] Dr. Vijayarangan & Dr. N. Ganesan, “Stress Analysis of Composite Gear using Finite Element Approach”, Computers and Structures(ISSN 0045- 7949) vol. 46, no. 5, p. 869- 875, March 1993.

[7] Dr. Charles Cooper, “Issues of Gear Design using 3 - D Solid ModelingSystems”, Randall Publications Inc., 2001.

[8] Dr. B. Pradhan, “Stress Analysis of Gears”, Summer School on Gear Engineering, I. S. T. E., May 1978.

[9] Dr. G. L. Sinha, “Geometry of Gear Teeth”, Summer School on Gear Engineering, I. S. T. E., May 1978.

[10] Dr. S. S. N. Murthy, “Gear Accuracy Standards”, Summer School on Gear Engineering, I. S. T. E., May 1978.

Design and Analysis of Driven Shaft

MUSHAM SRAVAN KUMAR

sravanmusham88@gmail.com

M Tech(CAD/CAM)

Mechanical Engineering Department

Malla Reddy College Of Engineering, Mysammaguda.

Hyderabad

ABSTRACT:- A shaft-driven bicycle is a bicycle that uses a drive shaft instead of a chain to transmit power from the pedals to the wheel arrangement. Shaft drives were introduced over a century ago, but were mostly supplanted by chain-driven bicycles due to the gear ranges possible with sprockets and derailleur. Recently, due to advancements in internal gear technology, a small number of modern shaft-driven bicycles have been introduced. Shaft-driven bikes have a large bevel gear where a conventional bike would have its chain ring. This meshes with another bevel gear mounted on the drive shaft.

Key words:- Ansys, Catia, Analysis of Driven Shaft

1.INTRODUCTION

A shaft-driven bicycle is a bicycle that uses a drive shaft instead of a chain to transmit power from the pedals to the wheel. Shaft drives were introduced over a century ago, but were mostly supplanted by chain-driven bicycles due to the gear ranges possible with sprockets and derailleur's. Recently, due to advancements in internal gear technology, a small number of modern shaft-driven bicycles have been introduced.

Shaft-driven bikes have a large bevel gear where a conventional bike would have its chain ring. This meshes with another bevel gear mounted on the drive shaft. The use of bevel gears allows the axis of the drive torque from the pedals to be turned through 90 degrees. The drive shaft then has another bevel gear near the rear wheel hub which meshes with a bevel gear on the hub where the rear sprocket would be on a conventional bike, and canceling out the first drive torque change of axis.

The 90-degree change of the drive plane that occurs at the bottom bracket and again at the rear hub uses bevel gears for the most efficient performance, though other mechanisms could be used, e.g. hobson's joints, worm gears or crossed helical gears.

The drive shaft is often mated to a hub gear which is an internal gear system housed inside the rear hub. Manufacturers of internal hubs suitable for use with shaft drive systems include NuVinci, Rohloff, Shimano, SRAM, and Sturmey-Archer.

1.1 Applications

The bevel gear has many diverse applications such as locomotives, [1]marine applications, automobiles, printing presses, cooling towers, power plants, steel plants, railway track inspection machines, etc.

For examples, see the following articles on:

- Bevel gears are used in differential drives, which can transmit power to two

axles spinning at different speeds, such as those on a cornering automobile.

- Bevel gears are used as the main mechanism for a hand drill. As the handle of the drill is turned in a vertical direction, the bevel gears change the rotation of the chuck to a horizontal rotation. The bevel gears in a hand drill have the added advantage of increasing the speed of rotation of the chuck and this makes it possible to drill a range of materials.

- The gears in a bevel gear planer permit minor adjustment during assembly and allow for some displacement due to deflection under operating loads without concentrating the load on the end of the tooth.

- Spiral bevelgears are important components on rotorcraft drive systems. These components are required to operate at high speeds, high loads, and for a large number of load cycles. In this application, spiral bevel gears are used to redirect the shaft from the horizontal gas turbine engine to the vertical rotor.

1.2 Driven Shaft

A shaft-driven bicycle is a bicycle that uses a drive shaft instead of a chain to transmit power from the pedals to the wheel. Shaft drives were introduced over a century ago, but were mostly supplanted by chain-driven bicycles due to the gear ranges possible with sprockets and derailleurs. Recently, due to advancements in internal gear technology, a small number of modern shaft-driven bicycles have been introduced.

Fig:1 Drive shaft housing

2. Automotive drive shaft Vehicles

An automobile may use a longitudinal shaft to deliver power from an engine/transmission to the other end of the vehicle before it goes to the wheels. A pair of short drive shafts is commonly used to send power from a central differential, transmission, or transaxle to the wheels.

2.1 Front-engine, rear-wheel drive

In front-engined, rear-drive vehicles, a longer drive shaft is also required to send power the length of the vehicle. Two forms dominate: The torque[2] tube with a single universal joint and the more common Hotchkiss drive with two or more joints. This system became known asSystème Panhard after the automobilecompany PanhardetLevassor patent ed it.Most of these vehicles have a clutch and gearbox (or transmission) mounted directly on the engine with a drive shaft leading to a final drive in the rear axle. When the vehicle is stationary, the drive shaft does not rotate. A few, mostly sports, cars seeking improved weight balance between front and rear, and most commonly Alfa Romeos or Porsche 924s, have instead used a rear-mounted transaxle. This places the clutch and transmission at the rear of the car and the drive shaft between them and the engine. In this case the drive shaft rotates continuously as long

as the engine does, even when the car is stationary and out of gear.

A drive shaft connecting a rear differential to a rear wheel may be called a half-shaft. The name derives from the fact that two such shafts are required to form one rear axle.

Early automobiles often used chain drive or belt drive mechanisms rather than a drive shaft. Some used electrical generators and motors to transmit power to the wheels.

Fig:2.1 A truck double propeller shaft

2.2 Front-wheel drive

In British English, the term "drive shaft" is restricted to a transverse shaft[6] that transmits power to the wheels, especially the front wheels. A drive shaft connecting the gearbox to a rear differential is called a propeller shaft, or prop-shaft. A prop-shaft assembly consists of a propeller shaft, a slip joint and one or more universal joints. Where the engine and axles are separated from each other, as on four-wheel drive and rear-wheel drive vehicles, it is the propeller shaft that serves to transmit the drive force generated by the engine to the axles.

Several different types of drive shaft are used in the automotive industry:

- One-piece drive shaft
- Two-piece drive shaft
- Slip-in-tube drive shaft

The slip-in-tube drive shaft is a new type that improves crash safety. It can be compressed to absorb energy in the event of a crash, so is also known as a collapsible drive shaft.

2.3 Motorcycle drive shaft

Drive shafts have been used on motorcycles since before WW1, such as the Belgian FN motorcycle from 1903 and the Stuart TurnerStellar motorcycle of 1912. As an alternative to chain and belt drives, drive shafts offer relatively maintenance-free operation, long life and cleanliness. A disadvantage of shaft drive on a motorcycle is that helical gearing, spiral bevel gearing or similar is needed to turn the power 90° from the shaft to the rear wheel, losing some power in the process. On the other hand, it is easier to protect the shaft linkages and drive gears from dust, sand, and mud.

BMW has produced shaft drive motorcycles since 1923; and Moto Guzzi have built shaft-drive V-twins since the 1960s. The British company, Triumph and the major Japanese brands, Honda, Suzuki, Kawasaki and Yamaha, have produced shaft drive motorcycles. All geared models of the Vespa scooter produced to date have been shaft-driven.[citation needed] Vespa's automatic models, however, use a belt.

Motorcycle engines positioned such that the crankshaft is longitudinal and parallel to the frame are often used for shaft-driven motorcycles. This requires only one 90° turn in power transmission, rather than two. Bikes from Moto Guzzi and BMW, plus the Triumph Rocket III and Honda ST series all use this engine layout.

Motorcycles with shaft drive are subject to shaft effect where the chassis climbs when power is applied. This effect, which is the opposite of that exhibited by chain-drive motorcycles, is counteracted with systems such

as BMW's Paralever, Moto Guzzi's CARC and Kawasaki's Tetra Lever.

Fig:2.2 The exposed drive shaft on BMW's first motorcycle, the R32

2.4 Marine drive shaft

On a power-driven ship, the drive shaft, or propeller shaft, usually connects the transmission inside the vessel directly to the propeller, passing through a stuffing box or other seal at the point it exits the hull. There is also a thrust block[4], a bearing to resist the axial force of the propeller. As the rotating propeller pushes the vessel forward, any length of drive shaft between propeller and thrust block is subject to compression, and when going astern to tension. Except for the very smallest of boats, this force isn't taken on the gearbox or engine directly.

Cardan shafts are also often used in marine applications between the transmission and either a propeller gearbox or water jet.

Locomotive drive shaft

The Shay,[3] Climax and Heisler locomotives, all introduced in the late 19th century, used quill drives to couple power from a centrally mounted multi-cylinder engine to each of the trucks supporting the engine. On each of these geared steam locomotives, one end of each drive shaft was coupled to the driven truck through a universal joint while the other end was powered by the crankshaft,transmission or another truck through a second universal joint. A quill drive also has the ability to slide lengthways, effectively varying its length. This is required to allow the bogies to rotate when passing a curve.

Cardan shafts are used in some diesel locomotives (mainly diesel-hydraulics, such as British Rail Class 52) and some electric locomotives (e.g. British Rail Class 91). They are also widely used in diesel multiple units.

Fig:2.3 The rear drive shaft, crankshaft and front drive shaft of a Shay locomotive

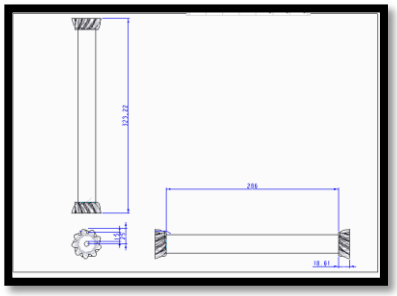
2.5 Working principle

The term[5] Drive shaft is used to refer to a shaft, which is used for the transfer of motion from one point to another. Whereas the shafts, which propel is referred to as the propeller shafts. However the drive shaft of the automobile is also referred to as the propeller shaft because apart from transmitting the rotary motion from the front end to the rear end of the vehicle, these shafts also propel the vehicle forward. The shaft is the primary connection between the front and the rear end, which performs both the jobs of transmitting the motion and propelling the front end.

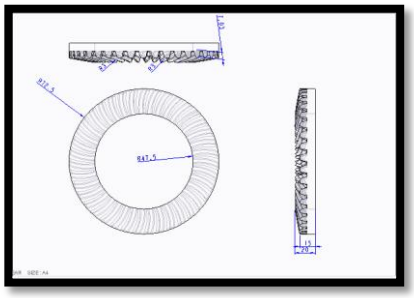
3. Design of Shaft

CATIA enables the creation of 3D parts, from 3D sketches, sheet metal, composites, and molded, forged or tooling parts up to the definition of mechanical assemblies. The software provides advanced technologies for mechanical surfacing. It provides tools to complete product definition, including functional tolerances as well as kinematics definition.

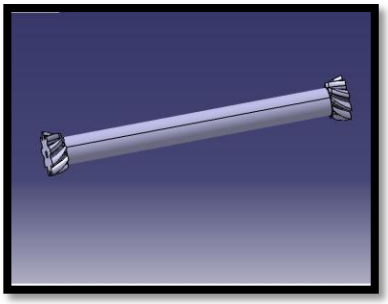
3.1 2D MODELS OF SHAFT



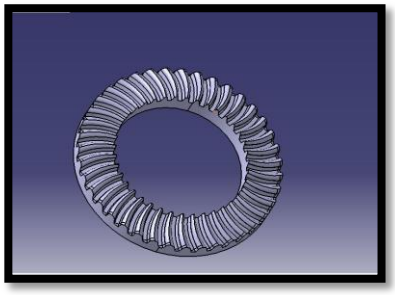
3.2 BEVEL GEAR



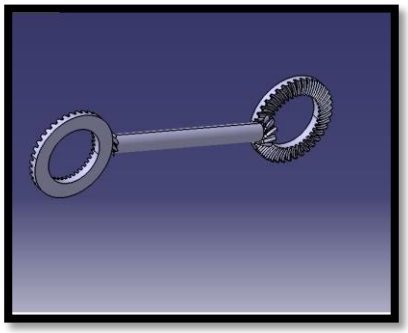
3.3 3D MODELS OF SHAFT



3.4 BEVEL GEAR



3.5 ASSEMBLY MODEL



4. Design Analysis

Assumptions the shaft rotates at a constant speed about its longitudinal axis. The shaft has a uniform, circular cross section. The shaft is perfectly balanced. Hexa Mesh is made for better result and 20000 elements made with fine mesh size. The regular FEA procedure is followed and obtained results were plotted and compared.

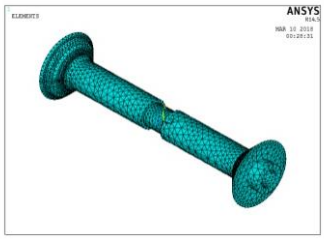


Fig:1 Boundary condition on driven shaft

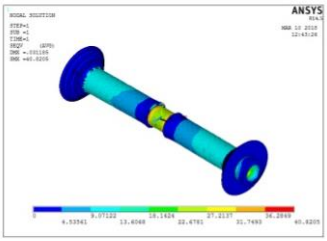


Fig:2 Stress formed on driven shaft

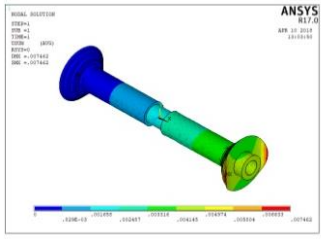


Fig:3 Resultunt displacment of driven shaft

4. CONCLUSION

Firstly the project were unable to be completed with the drive shaft due to various problems around circumference of the bicycle ,later on this was realized to run successfully with two bevel gears at both end of the drive shaft. The presented work was aimed to reduce the wastage of human power (energy) on bicycle riding or any machine, which employs drive shafts; in general it is achieved by using light weight drive shaft with bevel gears on both sides designed on replacing chain transmission. The presented work also deals with design optimization i.e converting rotary motion in linear motion with aid of two bevel gears. Instead of chain drive one piece drive shaft for rear wheel drive bicycle have been optimally designed and manufactured for easily power transmission. The drive shaft with the objective of minimization of weight of shaft which was subjected to the constraints such as torque transmission , torsion buckling capacity , stress, strain , etc The torque transmission capacity of the bicycle drive shaft has been calculated by neglecting and considering the effect of centrifugal forces and it has been observed that centrifugal force will reduce the torque transmission capacity of the shaft.The stress distribution and the maximum deformation in the drive shaft are the functions of the stacking of material. The optimum stacking of material layers can be used as the effective tool to reduce weight and stress acting on the drive shaft. The design of drive shaft is critical as it is subjected to combined loads. The designer has two options for designing the drive shaft whether to select solid or hollow shaft. The solid shaft gives a maximum value of torque transmission but at same time due to increase in weight of shaft, For a given weight, the hollow shaft is stronger because it has a bigger diameter due to less weight & less bending moment The results obtained from this work is an useful approximation to help in the earlier stages of the development, saving development time and helping in the decision making process to optimize a design. The drive shaft has served as an alternative to a chain-drive in bicycles for the past century, never becoming very popular

5. REFERENCES

1.Rastogi, N. (2004). Design of composite drive shafts for automotive applications. Visteon Corporation, SAE technical paper series.

2.Designs and Analysis of a Propeller Shaft of a Toyota Qualis by “Syed Hasan”.

3.Luther H. Porter, Wheels and Wheeling, Wheelman, Boston, 1892; page 162.

4.Jump up^ Walter Stillman, Bicycle, U.S. Patent 456,387, July 21, 1891.

5.Jump up^ *Archibald Sharp (1896).* Bicycles and Tricycles. *Longmans. p. 461.* ISBN 0-486-42987-3.

6.Jump up to:[a b c] Herlihy, David V. *(2004). Bicycle, the History. Yale University Press. pp. 286–287.*

HEAT TRANSFER ENHANCEMENT IN VERTICAL NARROW PLATES BY NATURAL CONVECTION

BIJJE SADANAND

sadanandbijje@gmail.com

M.TECH(THERMAL ENGINEERING)

MECHANICAL ENGINEERING DEPARTMENT

Malla Reddy College of Engineering,Mysammaguda

Hyderabad

ABSTRACT:- Natural Convection flow in a vertical channel with internal objects is encountered in several technological applications of particular interest of heat dissipation from electronic circuits, refrigerators, heat exchangers, nuclear reactors fuel elements, dry cooling towers, and home ventilation etc.In this thesis the air flow through vertical narrow plates is modeled using PRO-E design software. The thesis will focus on thermal and CFD analysis with different Reynolds number (2×10^6 & 4×10^6) and different angles (0^0,30^0,45^0&60^0) of the vertical narrow plates. Thermal analysis done for the vertical narrow plates by steel, aluminum & copper at different heat transfer coefficient values. These values are taken from CFD analysis at different Reynolds numbers.

Key words:-Pro E,Ansys, CFD Analysis

1.INTRODUCTION

1.1Natural Convection: In natural convection, the fluid motion occurs by natural means such as buoyancy. Since the fluid velocity associated with natural convection is relatively low, the heat transfer coefficient encountered in natural convection is also low.

Mechanisms of Natural Convection Consider a hot object exposed to cold air. The temperature of the outside of the object will drop as a result of heat transfer with cold air and the temperature of adjacent air to the object will rise. Consequently, the object is surrounded with a thin layer of warmer air and heat will be transferred from this layer to the outer layers of air.

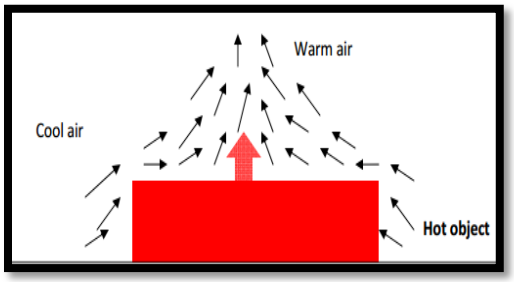


Fig:1Natural convection heat transfer from a hot body

The temperature of the air adjacent to the hot object is higher, thus its density is lower. As a resut, the heated air rises. This movement is called the natural convection current. Note that in the absence of this movement, heat transfer would be by conduction only and its rate would be much lower. In a gravitational field, there is a net force that pushes a light fluid placed in a heavier fluid upwards. This force is called the buoyancy force.

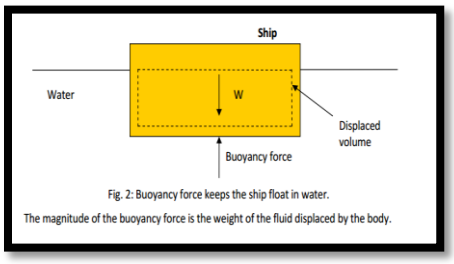


Fig:2 Natural Convection from a Vertical Plate

In this system heat is transferred from a vertical plate to a fluid moving parallel to it by natural convection. This will occur in any system wherein the density of the moving fluid varies with position. These phenomena will only be of significance when the moving fluid is minimally affected by forced convection.

2. LITERATURE REVIEW

In 1972, Aung et al. [12] presented a coupled numerical experimental study. Under isothermal conditions at high Rayleigh numbers their experimental results were 10% lower than the numerical ones. This difference has also been observed between Bodoia's and Osterle's numerical results [8] and Elenbaas' experimental ones [7]. They ascribed the discrepancies to the assumption of a flat velocity profile at the channel inlet.

However, the difference could also be attributed to the 2D hypothesis for the numerical simulations. In their 2D simulations in 1981, Dalbert et al. [13] introduced a pressure loss at the channel inlet in order to satisfy the Bernoulli equation between the hydrostatic conditions far from the channel and the channel inlet. Their results agreed better with the vertical flat plate regime than those of previous studies.

3. INTRODUCTION TO CAD

Throughout the history of our industrial society, many inventions have been patented and

whole new technologies have evolved. Perhaps the single development that has impacted

manufacturing more quickly and significantly than any previous technology is the digital computer. Computers are being used increasingly for both design and detailing of engineering components in the drawing office. Computer-aided design (CAD) is defined as the application of computers and graphics software to aid or enhance the product design from conceptualization to documentation. CAD is most commonly associated with the use of an interactive computer graphics system, referred to as a CAD system. Computer-aided design systems are powerful tools and in the mechanical design and geometric modeling of products and components.

There are several good reasons for using a CAD system to support the engineering design

Function:

- To increase the productivity
- To improve the quality of the design
- To uniform design standards
- To create a manufacturing data base
- To eliminate inaccuracies caused by hand-copying of drawings and inconsistency between
- Drawings

3.1 INTRODUCTION TO PRO/ENGINEER

Pro/ENGINEER, PTC's parametric, integrated 3D CAD/CAM/CAE solution, is used by discrete manufacturers for

mechanical engineering, design and manufacturing.Created by Dr. Samuel P. Geisberg in the mid-1980s, Pro/ENGINEER was the industry's first successful parametric, 3D CAD modeling system. The parametric modeling approach uses parameters, dimensions, features, and relationships to capture intended product behavior and create a recipe which enables design automation and the optimization of design and product development processes.

This powerful and rich design approach is used by companies whose product strategy is family-based or platform-driven, where a prescriptive design strategy is critical to the success of the design process by embedding engineering constraints and relationships to quickly optimize the design, or where the resulting geometry may be complex or based upon equations. Pro/ENGINEER provides a complete set of design, analysis and manufacturing capabilities on one, integral, scalable platform. These capabilities, include Solid Modeling, Surfacing, Rendering, Data Interoperability, Routed Systems Design, Simulation, Tolerance Analysis, and NC and Tooling Design.

Companies use Pro/ENGINEER to create a complete 3D digital model of their products. The models consist of 2D and 3D solid model data which can also be used downstream in finite element analysis, rapid prototyping, tooling design, and CNC manufacturing. All data is associative and interchangeable between the CAD, CAE and CAM modules without conversion. A product and its entire bill of materials(BOM) can be modeled accurately with fully associative engineering drawings, and revision control information. The associativity in Pro/ENGINEER enables users to make changes in the design at any time during the product development process and automatically update downstream deliverables. This capability enables concurrent engineering — design, analysis and manufacturing engineers working in parallel — and streamlines product development processes.

4. ANSYS Software:

ANSYS is an Engineering Simulation Software (computer aided Engineering). Its tools cover Thermal, Static, Dynamic, and Fatigue finite element analysis along with other tools all designed to help with the development of the product. The company was founded in 1970 by Dr. John A. Swanson as Swanson Analysis Systems, Inc. SASI. Its primary purpose was to develop and market finite element analysis software for structural physics that could simulate static (stationary), dynamic (moving) and heat transfer (thermal) problems. SASI developed its business in parallel with the growth in computer technology and engineering needs. The company grew by 10 percent to 20 percent each year, and in 1994 it was sold. The new owners took SASI's leading software, called ANSYS®, as their flagship product and designated ANSYS, Inc. as the new company name.

4.1 Benefits of ANSYS:

- The ANSYS advantage and benefits of using a modular simulation system in the design process are well documented. According to studies performed by the Aberdeen Group, best-in-class companies perform more simulations earlier. As a leader in virtual prototyping, ANSYS is unmatched in terms of

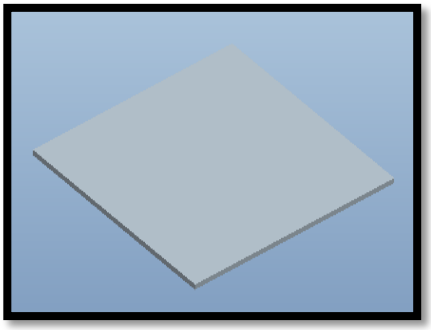
functionality and power necessary to optimize components and systems.

- The ANSYS advantage is well-documented.
- ANSYS is a virtual prototyping and modular simulation system that is easy to use and extends to meet customer needs; making it a low-risk investment that can expand as value is demonstrated within a company. It is scalable to all levels of the organization, degrees of analysis complexity, and stages of product development.

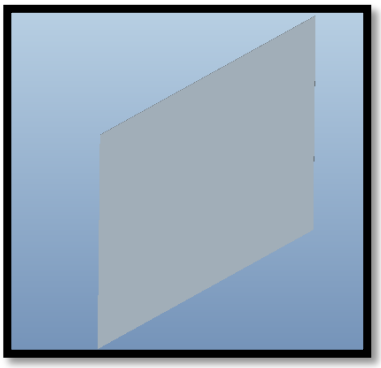
4.2 MODELLING AND ANALYSIS

The vertical narrow plate is modeled using the given specifications and design formula from data book. The isometric view of vertical narrow plate is shown in below figure. The vertical narrow plate profile is sketched in sketcher and then it is extruded vertical narrow plate using extrude option.

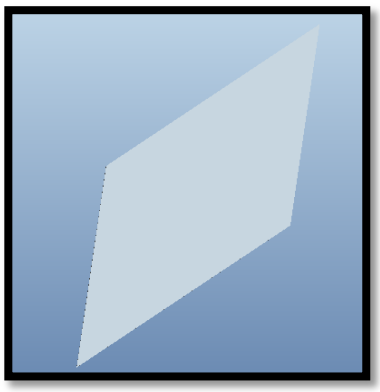
Vertical narrow plate at 0^0 3D model vertical narrow plates at 0^0 2D models



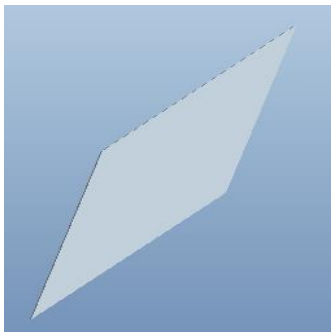
Vertical narrow plate at 30^0 3D model vertical narrow plates at 30^0 2D models

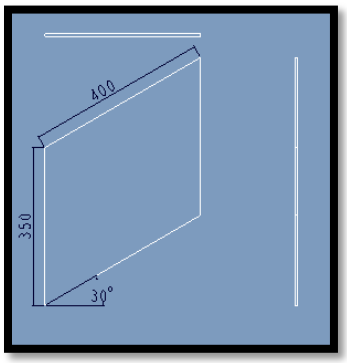


Vertical narrow plate at 45^0 3D model vertical narrow plates at 45^0 2D models



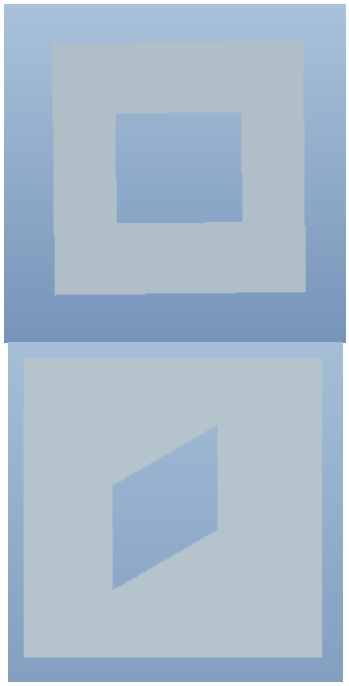
Vertical narrow plate at 60^0 3D model vertical narrow plates at 60^0 2D models



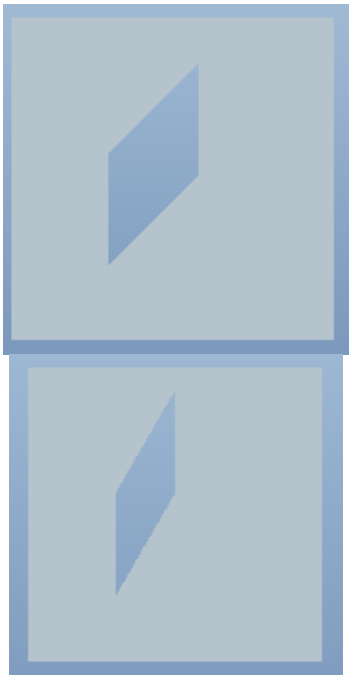


4.3 VERTICAL NARROW PLATE SURFACE MODELS

Vertical narrow plate at 0^0 3D models Vertical narrow plate at 30^0 3D models



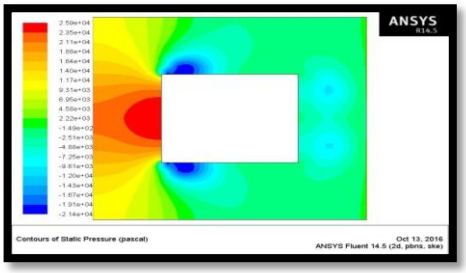
Vertical narrow plate at 45^0 3D models Vertical narrow plate at 60^0 3D models



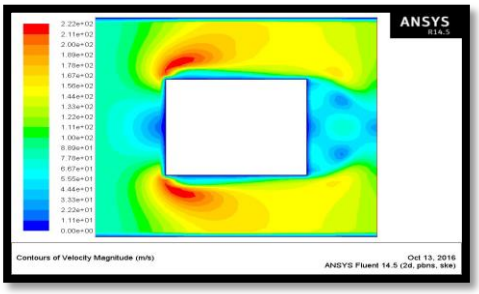
4.4 VERTICAL NARROW PLATE AT 0^0

REYNOLDS NUMBER - 2×10^6

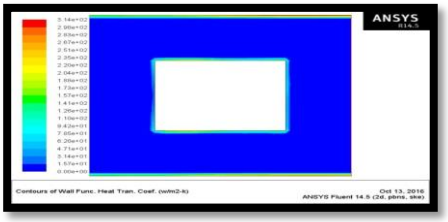
STATIC PRESSURE



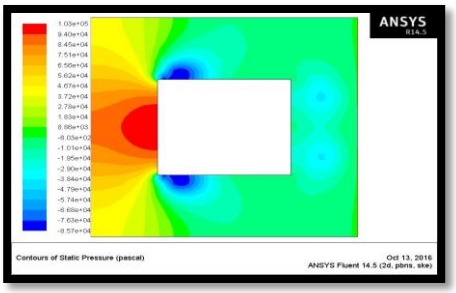
VELOCITY



HEAT TRANSFER COEFFICIENT

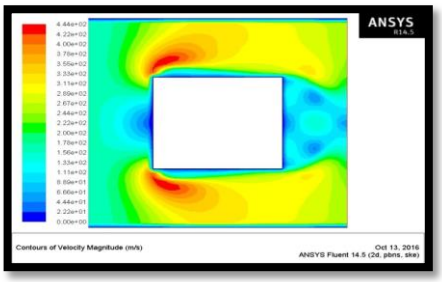


STATIC PRESSURE



.

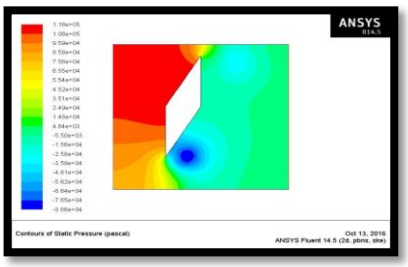
VELOCITY



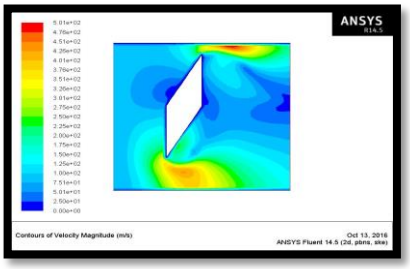
4.5 VERTICAL NARROW PLATE AT 60^0

REYNOLDS NUMBER - 2×10^6

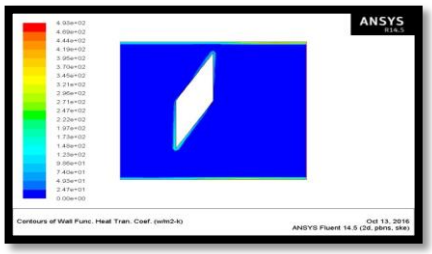
STATIC PRESSURE



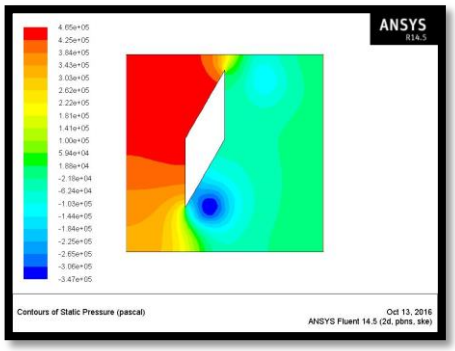
VELOCITY



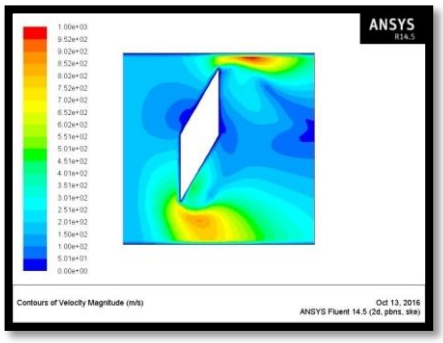
HEAT TRANSFER COEFFICIENT



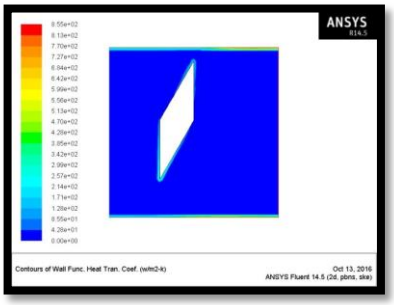
STATIC PRESSURE



VELOCITY



HEAT TRANSFER COEFFICIENT



5. CONCLUSION:

In this thesis the air flow through vertical narrow plates is modeled using PRO-E design software. The thesis will focus on thermal and CFD analysis with different Reynolds number (2×10^6 & 4×10^6) and different angles (0^0,30^0,45^0&60^0) of the vertical narrow plates. Thermal analysis done for the vertical narrow plates by steel, aluminum & copper at different heat transfer coefficient values. These values are taken from CFD analysis at different Reynolds numbers.

By observing the CFD analysis the pressure drop & velocity increases by increasing the inlet Reynolds numbers and increasing the plate angles. The heat transfer rate increasing the inlet Reynolds numbers, more heat transfer rate at 0^0 angles.

By observing the thermal analysis, the taken different heat transfer coefficient values are from CFD analysis. Heat flux value is more for copper material than steel& aluminum.

So we can conclude the copper material is better for vertical narrow plates.

6. REFERENCES

1.Arpaci, V. S., 1995, "Buoyant Turbulent Flow Driven by Internal EnergyGeneration," Int. J. Heat Mass Transfer, Vol. 38, pp.

2.761-2770.Asfia, F. J., and Dhir, V. K., 1994, "An 2.Experimental Study ofNarural Convectionin a Volumetrically Heated Spherical Pool With Rigid Wall," ASME Paper 94-WA/H7--26.

3.Cheung, F. B., i980a, "HeatSource-Driven Thermal Convection at Arbirrary Prandtl Numbers," J. Fluid Mech., YoL 97, pp. 743-758.

4. Cheung, F. B., 1980b, "The Boundry Layer Behavior in Transient Turbulent Thermal Convection Flow," ASME JoURNAL oF HEAr TRANSFERV, ol. 102, pp.373-375.

5.Cheung, F. B., 1978, "Turbulent Thermal Convection in a Horizontal FluidLayer With Time Dependent Volumetric Energy Sources." AIAA/ASME Thermo physics and Heat Transfer Conf.,78-HT-6, Palo Alto.

6. Cheung, F.8., 1977, "Natural Convection in a Volumerrically Heared FluidLayer at High Rayleigh Numbers," Int. J. Heat Mass Transfer, Vol. 20, pp.499-506.

7.Cheung, F. B., Shiah, S. W., Cho, D. H., and Tan, M. J., 1992, "Modeling ofHeat Transfer in a Horizontal HearGenerating Layer by an Effective DiffusivityApproach," ASME/HTD, Vol. 192, pp.55-62.

8.Dinh, T. N., and Nourgalier, R. R., 1997, "On Turbulence Modeling in LargeVolumetrically Heated Liquid Pools," Nncl. Engng. Design, in press.Fan, T. H., 1996, "Heat Transport Phenomena ofTurbulent Natural Convectionin a Melt Layer With Solidification," M.S. thesis, The Pennsylvania State University, University Park, PA.

9.Fielder, H. E., and Wille, R., 1970, "Turbulante Freie Konvektion in EinerHorizontalea Flussigkeitss chicht mit Volumen-Warmequelle," Paper NC 4.5,Proc. Fourth Int. Heat Transfer Conf., Vol. IV, pp. l-12.

DESIGN EVALUTION AND OPTIMIZATION OF NOZZLE USED IN DIESEL ENGINE

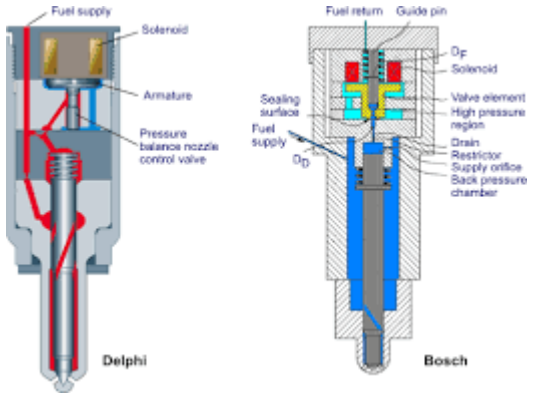
AREBOINA SANDEEP
areboinasandeep@gmail.com
M Tech(CAD/CAM)
Mechanical Engineering Department
Malla Reddy College Of Engineering , Mysammaguda.
Hyderabad

Abstract —
The nozzle is used to convert the chemical thermal energy generated in the combustion chamber into kinetic energy. The nozzle converts the low velocity, high pressure, high temperature gas in the combustion chamber into high velocity gas of lower pressure and temperature. Nozzle is a device designed to control the rate of flow, speed, direction, mass, shape, and/or the pressure of the stream that exhaust from them. Nozzles come in a variety of shapes and sizes depending on the mission of the rocket, this is very important for the understanding of the performance characteristics of rocket. Convergent divergent nozzle is the most commonly used nozzle since in using it the propellant can be heated in combustion chamber. In this thesis the convergent divergent nozzle changing the different nozzle diameters and different fluids at different velocities. We modeled convergent divergent nozzle changing with different nozzle diameters and Analyzed the convergent divergent nozzle with different mass flow rates to determine the pressure drop, heat transfer coefficient, and velocity and heat transfer rate for the fluid by CFD technique.

Keywords — ANSYS, Catia, optimization of nozzle.

1.INTRODUCTION

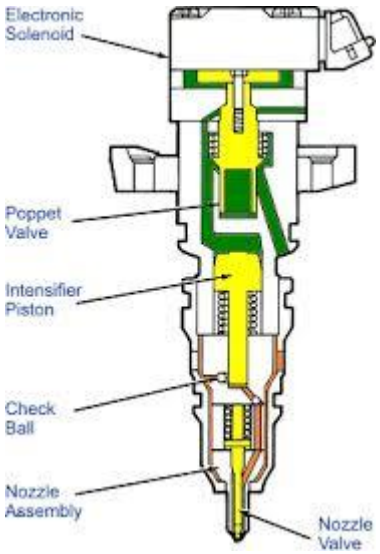
The primary challenges towards developing new diesel engines for passenger cars lie in the strict future emission legislation in combination with the customer's demands for steadily 0improving performance. For example, the emission limitations of Tier 2 Bin 5 requires an advanced after treatment system and a robust combustion process that minimizes emissions in the process of them being formed. Advancements in the technology of Diesel Injection (DI) systems have played in important role in the improvements that have been made up to this point . Combining the reduction in nozzle orifice diameters through enhanced flow characteristics with increased injection pressures provides an opportunity to develop engines featuring high power density and reduced emissions. The primary drawback to these modern spray hole geometries is that they often suffer a reduction of power output during long term operation. Other studies have identified these critical formations of deposits as the main reason for this behavior.



Basic mechanisms can be used to explain the formation and removal of deposits in internal combustion engines These mechanisms act independently of the location of formed deposits (e.g. injection

nozzles, heat changer) and of the combustion process (e.g. IDI, DI; diesel or gasoline).

The model described in the study illustrates the interaction of a wall with the enclosing flow regime. The transport of particles to the wall is based on the process of thermophores is This process results in the force of gas particles in the direction of the temperature depression. It is amplified with an increasing temperature differential between wall (cold)and fluid (hot). This process results is an increasing concentration of deposit-building particles near the wall.



High 2 turbulence near the wall may reduce the force of the aerosol again to a mean value, compensating for an increased temperature difference. The deposits are composed of attached particles (solid and liquid) and gas (Figure 1).

Condensation and adsorption of gaseous compounds at the cold wall promotes the formation process. At this point, the growth of the deposits is now mainly influenced by the sticking, impaction and incorporation of particles The adsorption of gaseous components and the chemical reactions (as pyrolysis, dehydration and polymerization, etc.), lead to the compaction of the deposits]. The removal of deposits has analogous physical mechanisms.

The chemical mechanism is oxidation destroying the organic compounds in the coating Evaporation and desorption reduce the gaseous fraction dissolved in the deposits. Abrasion is caused by strong aerodynamic forces and breaking-off, due to high temperature changes, resulting in inhomogeneous extensions of the wall and deposit layer.

The corresponding shearing stresses initiate the breaking-off process The soluble fraction of the deposits is washed off by solvents (e.g. water as solvent for salt compounds)

2.LITERATURE REVIEW

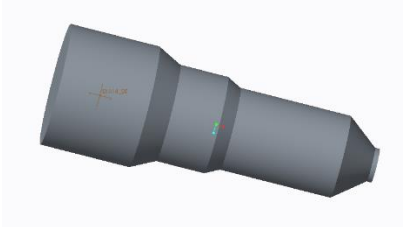
Design and Optimization of Fuel Injection System in Diesel Engine Using Biodiesel – A Review H. M. Pate

Fuel injection is systems for supplying high pressurize fuel to maximum mixing of fuel with air in an internal combustion engine. Direct Injection (DI) Systems as used in DI engines, in which the fuel is injected directly into a combustion chamber formed in the clylinder itself. The fuel injector directly injects fuel into the direct fuel injection system. The injector is a very complicated part, and massive research has been done to improve it.In my work indicating the development of fuel injector system to reduce chocking problem which is generally happen in bio diesel engine. The injection nozzles and their respective nozzle holders are vitally important components situated between the in-line injection pump and the diesel engine, its functions are as metering the injection of fuel, management of the fuel, defining the rate-of-discharge curve, Sealing-off against the combustion, chamber. Mechanical type injectors used in direct injection system. When biodiesel is used in the diesel engine choking problem is created in fuel injector. Therefore, we optimize the design of fuel injector component, and tried to prevent the chocking problem. The diesel fuel injector system directly injects fuel into the system without chocking.

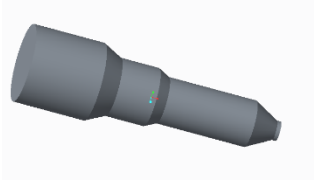
M. Volmajer et al [4] had numerical and experimental results of the nozzle fuel flow analysis for a four-hole injection nozzle Bosch DLLA 148 S 311376 are presented. The fuel flow coefficients obtained from the experimental results at steady flow conditions in the nozzle are compared with the results of the CFD analysis. The fuel flow coefficients obtained from the experimental results at steady flow conditions in the nozzle are compared with the results of the CFD analysis. From the presented results the following conclusions could be made. Flow coefficient testing device constructed at the ERL yields sufficiently precision, with reasonable uncertainties of the measurement. To refine the precision of the measurement, by defining the exact value of the pressure difference, the pressure downstream of the nozzle should be measured, or the nozzle position should be changed so, that the fluid would be injected directly into the measuring Plexiglas. For the same purpose, Plexiglas cylinder with high ovalness should be replaced with the glasslPlexiglas cylinder with proper circle cross-section. the presented testing device also enables the measurement of the flow coefficient separately for each nozzle hole, which brings better comparison with the results of CFD analysis when the simplified models, introducing only one hole, are applied. Zhijun Li et al [5]had investigates the effects of manufacturing variations in fuel injectors on the engine performance with emphasis on emissions. The variations are taken into consideration within a Reliability-Based Design Optimization (RBDO) framework. A reduced version of Multi-Zone Diesel engine Simulation (MZDS), MZDS-lite, is used to enable the optimization study. The numerical noise of MZDS-lite prohibits the use of gradient-based optimization methods. Therefore, surrogate models are developed to filter out the noise and to reduce computational cost.

Three multi-objective optimization problems are formulated, solved and compared: deterministic optimization using MZDS-lite, deterministic optimization using surrogate models and RBDO using surrogate models. The obtained results confirm that manufacturing variation effects must be taken into account in the early product development stages. The effects of manufacturing variations in fuel injectors on the engine performance with emphasis on emissions. The results obtained using deterministic and probabilistic optimization formulations demonstrated the need for RBDO to improve not only performance but also reliability. LI Minghai et al [7] had indicated forced lubrication is adopted for the new injector nozzle matching parts, which can reduce failure rate and increase service life. If the patented product is used widely, economic efficiency and social efficiency will be obtained. Benny Paul et al [8] had indicated effect of helical, spiral, and helical-spiral combination manifold configuration on air motion and turbulence inside the cylinder. Swirl inside the engine is important for diesel engine. Hence, for better performance they recommended a helical-spiral inlet manifold configuration.

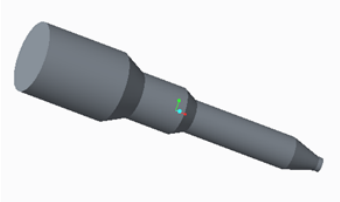
2.1 3D MODEL OF DIESEL NOZZLE WITH 50DIA



2.2 NOZZLE WITH 40DIA



2.3 NOZZLE WITH 30DIA



3. INTRODUCTION TO FEA

Finite element analysis is a method of solving, usually approximately, certain problems in engineering and science. It is used mainly for problems for which no exact solution, expressible in some mathematical form, is available. As such, it is a numerical rather than an analytical method. Methods of this type are needed because analytical methods cannot cope with the real, complicated problems that are met with in engineering. For example, engineering strength of materials or the mathematical theory of elasticity can be used to calculate analytically the stresses and strains in a bent beam, but neither will be very successful in finding out what is happening in part of a car suspension system during cornering.

One of the first applications of FEA was, indeed, to find the stresses and strains in engineering components under load. FEA, when applied to any realistic model of an engineering component, requires an enormous amount of computation and the development of the method has depended on the availability of suitable digital computers for it to run on. The method is now applied to problems involving a wide range of phenomena, including vibrations, heat conduction, fluid mechanics and electrostatics, and a wide range of material properties, such as linear-elastic (Hookean) behavior and behavior involving deviation from Hooke's law (for example, plasticity or rubber-elasticity).

Many comprehensive general-purpose computer packages are now available that can deal with a wide range of phenomena, together with more specialized packages for particular applications, for example, for the study of dynamic phenomena or large-scale plastic flow. Depending on the type and complexity of the analysis, such packages may run on a microcomputer or, at the other extreme, on a supercomputer. FEA is essentially a piece-wise process. It can be applied to one-dimensional problems, but more usually there is an area or volume within which the solution is required. This is split up into a number of smaller areas or volumes, which are called finite elements. Figure 1 shows a two-dimensional model of a spanner that has been so divided: the process is called discretisation, and the assembly of elements is called a mesh.

4.INTRODUCTION TO ANSYS

4.1 Structural Analysis

ANSYS Autodyn is computer simulation tool for simulating the response of materials to short duration severe loadings from impact, high pressure or explosions.

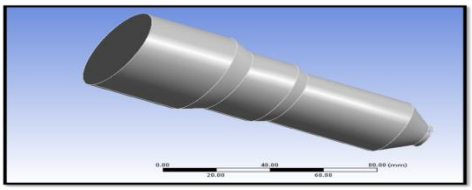
4.2 Fluid Dynamics

ANSYS Fluent, CFD, CFX, FENSAP-ICE and related software are Computational Fluid Dynamics software tools used by engineers for design and analysis. These tools can simulate fluid flows in a virtual environment — for example, the fluid dynamics of ship hulls; gas turbine engines (including the compressors, combustion chamber, turbines and afterburners); aircraft aerodynamics; pumps, fans, HVAC systems, mixing vessels, hydro cyclones, vacuum cleaners, etc.

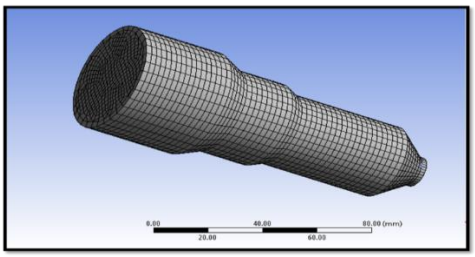
4.3 CFD ANALYSIS OF DIESEL ENGINE NOZZLE

FLUID- DIESEL

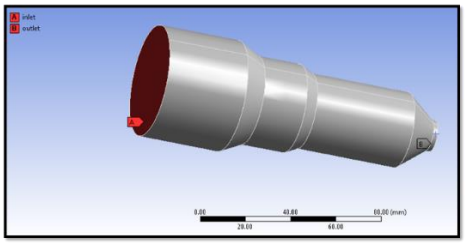
Velocity inlet = 200m/s, 300m/s & 400m/s



MESHED MODEL



SPECIFYING THE BOUNDARIES FOR INLET & OUTLET

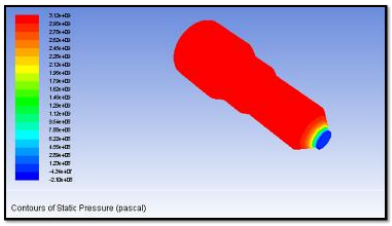


4.4 FLUID- DIESEL

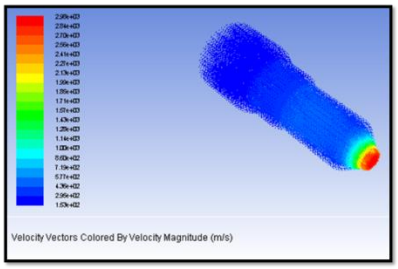
DIESEL ENGINE NOZZLE DIA. 50MM

VELOCITY INLET = 200m/s

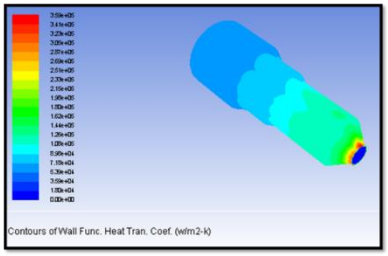
PRESSURE



VELOCITY

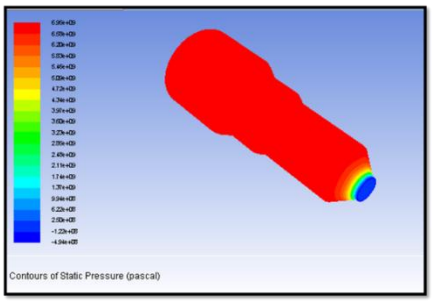


HEAT TRANSFER COEFFICIENT

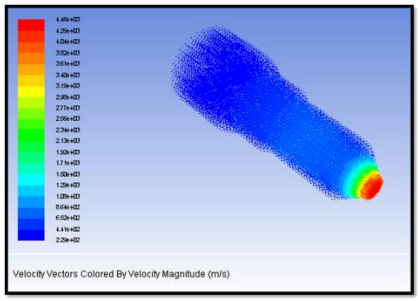


4.5 VELOCITY INLET = 300m/s

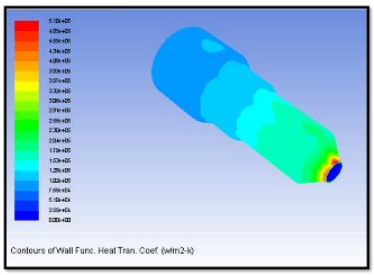
PRESSURE



VELOCITY



HEAT TRANSFER COEFFICIENT

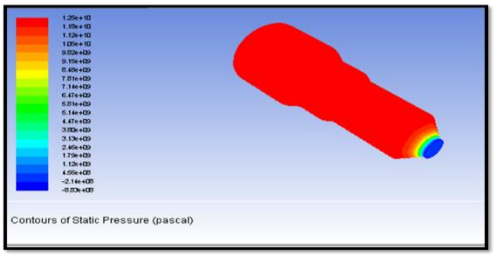


MASS FLOW RATE & HEAT TRANSFER RATE

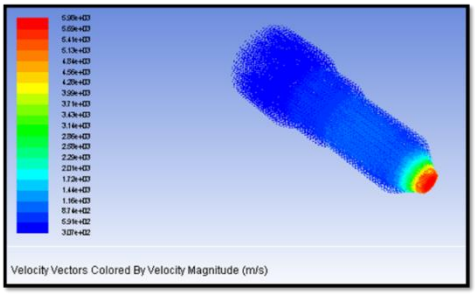
Mass Flow Rate	(kg/s)
inlet	427.56219
interior-____msbr	26587.756
outlet	-427.85144
wall-____msbr	0
Net	-0.28924561
otal Heat Transfer Rate	**(w)**
inlet	4333986
outlet	-4336913.5
wall-____msbr	0
Net	-2927.5

4.6 VELOCITY INLET = 400m/s

PRESSURE



VELOCITY



5. CONCLUSION

Nozzles come in a variety of shapes and sizes depending on the mission of the rocket, this is very important for the understanding of the performance characteristics of rocket. Convergent divergent nozzle is the most commonly used nozzle since in using it the propellant can be heated in combustion chamber.

In this thesis the convergent divergent nozzle changing the different nozzle diameters and different fluids at different velocities. We modeled convergent divergent nozzle changing with different nozzle diameters.

By observing the cfd analysis of diesel engine nozzle the pressure, velocity, heat transfer rate and mass flow rate values are increases by increasing the inlet velocities and decreasing the nozzle dia.

By observing the thermal analysis, heat flux is more for aluminum alloy compared with brass material.

So it can be concluded the diesel engine nozzle efficiency were more when the nozzle dia. decreases.

6. REFERENCES

[1] J.B. Heywood, "Internal Combustion Engine Fundamentals",McGraw-Hill Book Co, pp 493-494, 1988.

[2] D. Ing. H. Tschöke, "Diesel distributor fuel-injection pumps", Robert Bosch GmbH, pp 12-53, 1999.

[3] B. Challen R. Baranescu, "Diesel Engine Reference Book" Reed Educational and Professional Publishing Ltd., Second Edition, pp.260-301, 1999.

[4] M. Volmajer, B. Kegl, "Experimental and numerical analysis of fuel flow in the diesel engine injection nozzle" , Journal of Kones. Combustion Engines, Vol. 8, No. 1-2, 2001.

[5] Z. Li, M. Kokkolaras, D. Jung, Panos Y. Papalambros and D. N. Assanis, "An Optimization Study of Manufacturing Variation Effects on Diesel Injector Design with Emphasis on Emissions", SAE International, 2004.

[6] A.J.Von Wielligh, "Influence of fuel quality on diesel injector failures", Fifth International Colloquium Fuels, Germany, 2005. Property Unit Jatropha oil Jatropha Oil methyl ester Diesel Density at 15°C kg/m3 918 880 850 Viscosity at 40°C mm2 /s 35.4 4.84 2.60 Flash point °C 186 162 70 International Journal of Innovations in Engineering and Technology (IJIET) Vol. 2 Issue 1 February 2013 76 ISSN: 2319 – 1058

[7] LI Minghai, CUI Hongjiang, W.Juan, G. Ying, "Improvement of fuel injection system of locomotive diesel engine", Journal of Environmental Sciences, pp. S139-S141, 2009.

[8] B. Paul, V. Ganesan, "Flow field development in a direct injection diesel engine with different manifolds" International Journal of Engineering, Science and Technology Vol. 2, No. 1, pp. 80-91, 2010.

[9] K. K.Khatri, D. Sharma, S. L. Soni, S. Kumar, D Tanwar, "Investigation of Optimum Fuel Injection Timing of Direct Injection CI Engine Operated on Preheated Karanj-Diesel Blend" Jordan Journal of Mechanical and Industrial Engineering Volume 4, pp. 629-640, 2010.

[10] Bizhan Befrui, Giovanni Corbinelli, Mario D'Onofrio and Daniel Varble, "GDI Multi-Hole Injector Internal Flow and Spray Analysis", SAE International, 2011.

DESIGN AND CFD ANALYSIS OF FLOW THROUGH VENTURI OF A CARBURETOR

PELIMALI RAVICHANDRA

Pelimeli09@gmail.com

M.TECH(THERMAL ENGINEERING)

MECHANICAL ENGINEERING DEPARTMENT

Malla Reddy College of Engineering,Mysammaguda

Hyderabad

ABSTRACT:*The process of forming a combustible fuel-air mixture by mixing the right amount of fuel with air before admission to the cylinder of the engine is called carburetion and the device doing this job is called carburetor.Modern passenger vehicles with gasoline engines are provided with different compensating devices for fuel air mixture supply. One of the important factors that affect the fuel consumption is that design of carburetor. The venture of the carburetor is important that provides a necessary pressure drop in the carburetor device. Since different SI engine alternative fuels such as LPG, CNG are used in the present day vehicles to reduce the pollution and fuel consumption. Still for a better economy and uniform fuel air supply there is a need to design the carburetor with an effective analytical tool or software.In this project venture of carburetor is modeled in 3D modeling software Creo/Engineer. CFD analysis is done on the venture by varying the fuel discharge nozzle angle on the flow.*

Key word: Ansys, Pro E,CFD Analysis

1.INTRODUCTION

Engine is a device that transforms one form of energy into another form. Heat energy is a device that transforms the chemical energy contained in a fuel to another form of energy and utilizes that energy for some useful work

1.1 Carburetor

The simple carburetor consists of the following basic parts

Float chamber

- Venturi
- Fuel discharge nozzle
- Metering orifice
- Choke
- Throttle valve

1.3 PRINCIPLE OF CARBURETION

Both air and gasoline are drawn into the cylinder due to suction pressure created by the downward movement of the piston. In the carburetor, the air passing into the combustion chamber picks up the fuel discharged by a fine orifice in a tube called the carburetor jet. The rate of discharge of the fuel depends on the pressure difference between the float chamber and the throat of the venturi of the carburetor and the area of the outlet of the tube. In order that the fuel is strongly atomized the suction effect must be strong and the nozzle outlet must be comparatively small. To produce a strong suction, a restriction is generally provided in the pipe in the carburetor carrying air to the engine. This restriction is called throat. In this throat due to increase in the velocity of the air the pressure is decreased and suction is created.

The venturi tube has a narrower path at the center so that the path through air is going to travel is reduced. As same amount of air must travel must travel through the

path of the tube so the velocity of the air at the venturi is increased and suction is created.

Usually the fuel discharge jet is located at the point where the suction is maximum. So this is positioned just below the throat of the venturi. The spray of the fuel from the fuel discharge jet and the air are mixed at this point of the throat and a combustible mixture is formed. Maximum amount of fuel gets atomized and some part gets vaporized. Due to increase in the velocity of the air at the throat the vaporization of the fuel becomes easier.

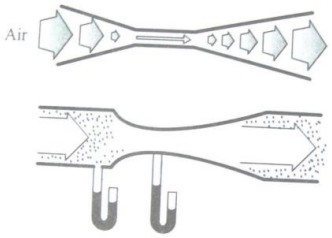


Fig:1 Operation of venture tube

2.PROJECT DESCRIPTION

In this project venture of carburetor is modeled in 3D modeling software Pro/Engineer. CFD analysis is done on the venture by varying the fuel discharge nozzle angle on the flow. The analysis was done for Θ = 30, 35, 40 and 45 where Θ is the angle between the axis of the fuel discharge nozzle and the vertical axis of the body of the carburetor.

And also CFD analysis is done to calculate the throat pressure for different angles of the throttle plate 45, 60, 75, 90.

Fig The simple carburetor

3.INTRODUCTION TO CREO

PTC CREO, formerly known as Pro/ENGINEER, is 3D modeling software used in mechanical engineering, design, manufacturing, and in CAD drafting service firms. It was one of the first 3D CAD modeling applications that used a rule-based parametric system. Using parameters, dimensions and features to capture the behavior of the product, it can optimize the development product as well as the design itself.

The name was changed in 2010 from Pro/ENGINEER Wildfire to CREO. It was announced by the company who developed it, Parametric Technology Company (PTC), during the launch of its suite of design products that includes applications such as assembly modeling, 2D orthographic views for technical drawing, finite element analysis and more.

PTC CREO says it can offer a more efficient design experience than other modeling software because of its unique features including the integration of parametric and direct modeling in one platform. The complete suite of applications spans the spectrum of product development, giving designers options to use in each step of the process. The software also has a more user friendly interface that provides a better experience for designers. It also has

collaborative capacities that make it easy to share designs and make changes.

There are countless benefits to using PTC CREO. We'll take a look at them in this two-part series.

First up, the biggest advantage is increased productivity because of its efficient and flexible design capabilities. It was designed to be easier to use and have features that allow for design processes to move more quickly, making a designer's productivity level increase.

Part of the reason productivity can be increased is because the package offers tools for all phases of development, from the beginning stages to the hands-on creation and manufacturing. Late stage changes are common in the design process, but PTC CREO can handle it. Changes can be made that are reflected in other parts of the process.

The collaborative capability of the software also makes it easier and faster to use. One of the reasons it can process information more quickly is because of the interface between MCAD and ECAD designs. Designs can be altered and highlighted between the electrical and mechanical designers working on the project.

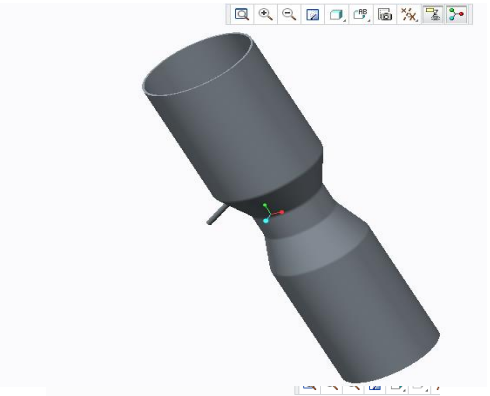
The time saved by using PTC CREO isn't the only advantage. It has many ways of saving costs. For instance, the cost of creating a new product can be lowered because the development process is shortened due to the automation of the generation of associative manufacturing and service deliverables.

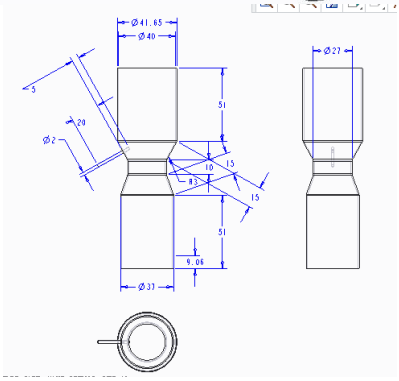
PTC also offers comprehensive training on how to use the software. This can save businesses by eliminating the need to hire new employees. Their training program is available online and in-person, but materials are available to access anytime.

A unique feature is that the software is available in 10 languages. PTC knows they have people from all over the world using their software, so they offer it in multiple languages so nearly anyone who wants to use it is able to do so.

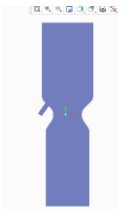
3.1 MODELS OF VENTURE

3D MODEL

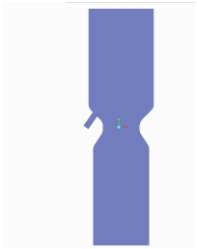




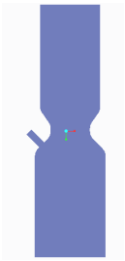
3.2 FUEL DISCHARGE NOZZLE ANGLE – 30^0



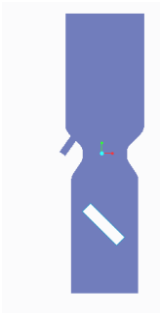
3.3 FUEL DISCHARGE NOZZLE ANGLE – 35^0



3.4 FUEL DISCHARGE NOZZLE ANGLE – 45^0



3.5 THROTTLE PLATE ANGLE – 45^0



.

4.INTRODUCTION TO ANSYS

4.1 Structural Analysis

ANSYS Autodyn is computer simulation tool for simulating the response of materials to short duration severe loadings from impact, high pressure or explosions.

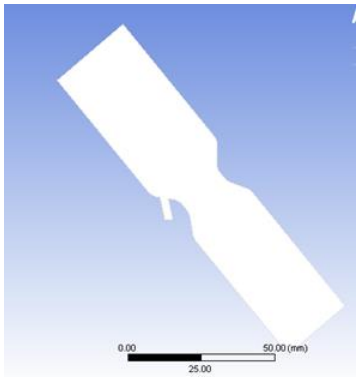
4.2 ANSYS Mechanical

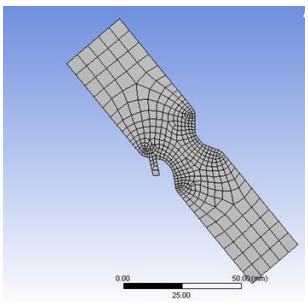
ANSYS Mechanical is a finite element analysis tool for structural analysis, including linear, nonlinear and dynamic studies. This computer simulation product provides finite elements to model behavior, and supports material models and equation solvers for a wide range of mechanical design problems. ANSYS Mechanical also includes thermal analysis and coupled-physics capabilities involving acoustics, piezoelectric, thermal–structural and thermo-electric analysis.

4.3 CFD ANALYSIS OF VENTURE OF CARBURETOR:

FUEL DISCHARGE NOZZLE ANGLE – 30^0:

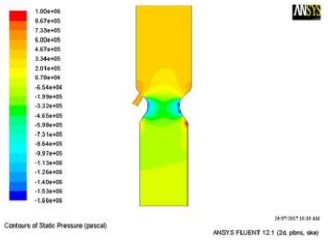
Save Creo- Model as .iges format



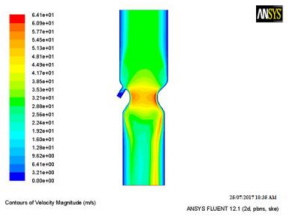


4.4 SPECIFYING BOUNDARIES FOR INLET AND OUTLET

Static Pressure

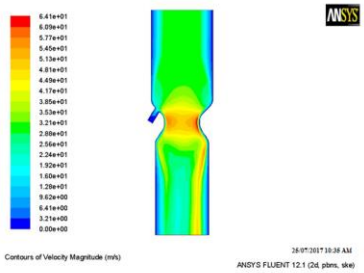


Velocity

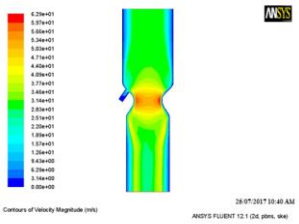


4.5 FUEL DISCHARGE NOZZLE ANGLE – 35^0

Static Pressure

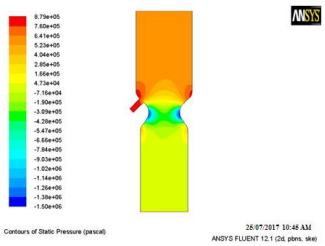


Velocity

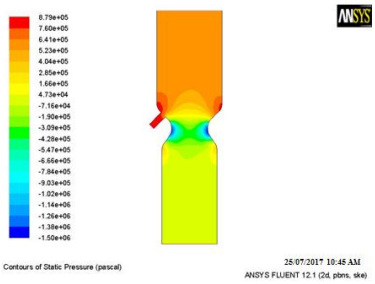


4.6 FUEL DISCHARGE NOZZLE ANGLE – 45^0

Static Pressure

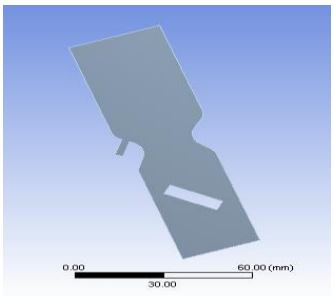


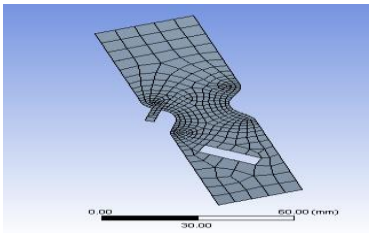
Velocity



4.7 THROTTLE PLATE ANGLE – 45^0

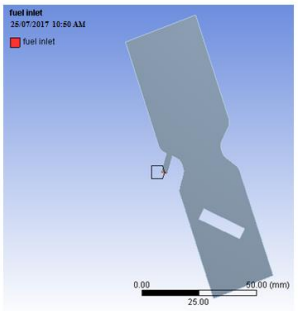
Save Creo Model as .iges format



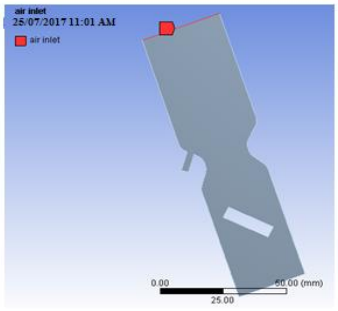


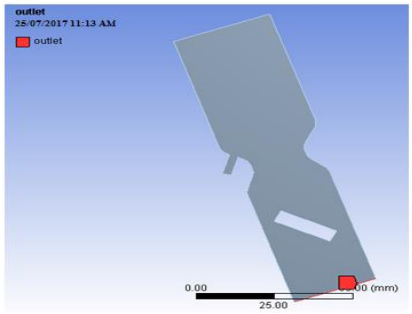
4.8 SPECIFYING BOUNDARIES FOR INLET AND OUTLET

Fuel Inlet

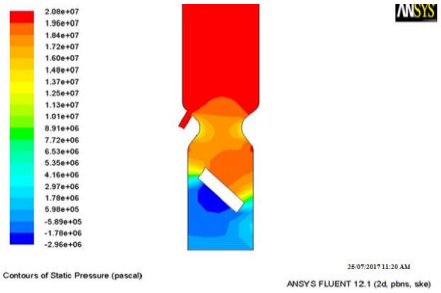


Air Inlet



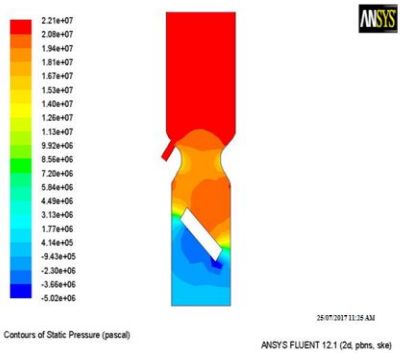


Static Pressure



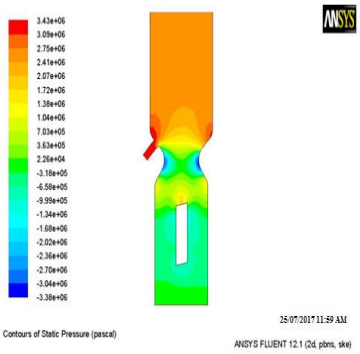
4.9 THROTTLE PLATE ANGLE – 60^0

Static Pressure



THROTTLE PLATE ANGLE – 90^0

Static Pressure



RESULTS TABLE

FUEL DISCHARGE ANGLE	30	35	40	45
STATIC PRESSURE (Pa)	$1e^6$	$6.75\ e^5$	$7.4\ e^5$	$8.79\ e^5$
VELOCITY (m/s)	$6.41\ e^1$	$6.21\ e^1$	$6.33\ e^1$	$6.37\ e^1$

THROTTLE PLATE ANGLE	45	60	75	90
STATIC PRESSURE (Pa)	2.08^7	$2.21\ e^7$	$4.22\ e^6$	$3.43\ e^6$

5. CONCLUSION

From the above analysis the conclusions obtained are

1. When the flow inside the carburetor was analyzed for different angles of throttle plate opening, it was found that the pressure at the throat of the venturi decreased with the increase in opening of the throttle plate. Because when the throttle plate opening increases then the flow of air through the carburetor increases but the fuel flow remains constant. So the mixture becomes leaner. But as obtained from the analysis above the pressure at the throat the throat also decreases with increase in opening of the throttle plate so the flow of fuel from the float chamber into the throat increases and hence the quality of the mixture tends to remain constant.

2. When analyzed for fuel discharge nozzle angle of 30^0, it was observed that the pressure distribution inside the body of the carburetor is quite uniform which leads to a better atomization and vaporization of the fuel inside the carburetor body. But in other cases like where the fuel discharge nozzle angle was 35^0, 40^0 or 45^0, the pressure distribution is quite non-uniform inside the body of the carburetor. So it is concluded that for gasoline operated engine the optimum fuel discharge nozzle angle is 30^0.

6. REFERENCES

[1] Ganeshan V. INTERNAL COMBUSTION ENGINES. New Delhi, TMH publication,2009

[2] Diego A. Arias. Numerical and experimental study of air and fuel flow in small engine carburetors. University of Wisconsin-Madison, 2005

[3] Heywood John B. Internal combustion engines fundamentals. Mc Graw Hill, Inc, 1988.

[4] Diego A. Arias, Timothy A. Shedd, Steady and non-steady flow in a simple carburetor, Inst Mech Eng (Lond) Proc , Volume 188, Issue 53, 1974, Pages 537-548.

[5] Fluent. FLUENT 5 User Guide, 1999.

[6] Gambit. GAMBIT 5 User Guide, 1999.

[7] Sayma Abdulnaser, Computational Fluid Dynamics, ISBN- 978-87-7681-938-4, Abdulnaser Sayma and Ventus Publishers, 2009

[8] D. L. Harrington. Analysis and digital simulation of carburetor metering. PhD thesis, University of Michigan, 1968.

[9] Martin Cook, Masud Behnia. Pressure drop calculation and modelling of inclined intermittent gas liquid flow, Chemical Engineering Science 55 (2000) page 4699-4708.

[10] Bhramara P., Rao V. D., Sharma K. V. , and Reddy T. K. K. CFD analysis of two phase flow in a horizontal pipe – prediction of pressure drop, International Journal of Mechanical,industrial and Aerospace Engineering 3:2 2009.

DESIGN AND ANALYSIS OF TURBINE WING

ALUGALA SRAVAN
sravanalugala6@gmail.com
M.TECH(CAD/CAM)
MECHANICAL ENGINEERING DEPARTMENT
Malla Reddy College of Engineering,Mysammaguda
Hyderabad

ABSTRACT:- A Turbine blade is the individual component which makes up the turbine section of a gas turbine. The blades are responsible for extracting energy from the high temperature, high pressure gas produced by the combustor. The turbine blades are often the limiting component of gas turbines. To survive in this difficult environment, turbine blades often use exotic materials like super alloys and many different methods of cooling, such as internal air channels, boundary layer cooling, and thermal barrier coatings.

Key words:-Analysis,Modeling,Design

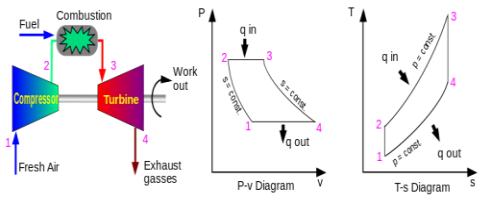
1.INTRODUCTION TO GAS TURBINE

A gas turbine, also called a combustion turbine, is a type of internal combustion engine. It has an upstream rotating compressor coupled to a downstream turbine, and a combustion chamber or area, called a combustor, in between.

The basic operation of the gas turbine is similar to that of the steam power plant except that air is used instead of water. Fresh atmospheric air flows through a compressor that brings it to higher pressure. Energy is then added by spraying fuel into the air and igniting it so the combustion generates a high-temperature flow. This high-temperature high-pressure gas enters a turbine, where it expands down to the exhaust pressure, producing a shaft work output in the process. The turbine shaft work is used to drive the compressor and other devices such as an electric generator that may be coupled to the shaft. The energy that is not used for shaft work comes out in the exhaust gases, so these have either a high temperature or a high velocity. The purpose of the gas turbine determines the design so that the most desirable energy form is maximized. Gas turbines are used to power aircraft, trains, ships, electrical generators, and tanks.

1.1 Theory of operation

In an ideal gas turbine, gases undergo four thermodynamic processes: an isentropic compression, isobaric (constant pressure) combustion, an isentropic expansion and heat rejection. Together, these make up the Brayton cycle.



In a real gas turbine, mechanical energy is changed irreversibly (due to internal friction and turbulence) into pressure and thermal energy when the gas is compressed (in either a centrifugal or axial compressor). Heat is added in the combustion chamber and the specific volume of the gas increases, accompanied by a slight loss in pressure. During expansion through the stator and rotor passages in the turbine, irreversible energy transformation once again occurs.

Fresh air is taken in, in place of the heat rejection.

If the engine has a power turbine added to drive an industrial generator or a helicopter rotor, the exit pressure will be as close to the entry pressure as possible with only enough energy left to overcome the pressure losses in the exhaust ducting and expel the exhaust. For a turboprop engine there will be a particular balance between propeller power and jet thrust which gives the most economical operation. In a jet engine only enough pressure and energy is extracted from the flow to drive the compressor and other components. The remaining high-pressure gases are accelerated to provide a jet to propel an aircraft.

The smaller the engine, the higher the rotation rate of the shaft(s) must be to attain the required blade tip speed. Blade-tip speed determines the maximum pressure ratios that can be obtained by the turbine and the compressor. This, in turn, limits the maximum power and efficiency that can be obtained by the engine. In order for tip speed to remain constant, if the diameter of a rotor is reduced by half, the rotational speed must double. For example, large jet engines operate around 10,000 rpm, while micro turbines spin as fast as 500,000 rpm.

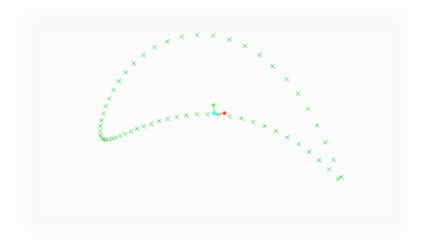
Mechanically, gas turbines can be considerably less complex than internal combustion piston engines. Simple turbines might have one main moving part, the compressor/shaft/turbine rotor assembly (see image above), with other moving parts in the fuel system. However, the precision manufacture required for components and the temperature resistant alloys necessary for high efficiency often make the construction of a simple gas turbine more complicated than a piston engine.

More advanced gas turbines (such as those found in modern jet engines) may have 2 or 3 shafts (spools), hundreds of compressor and turbine blades, movable stator blades, and extensive external tubing for fuel, oil and air systems.

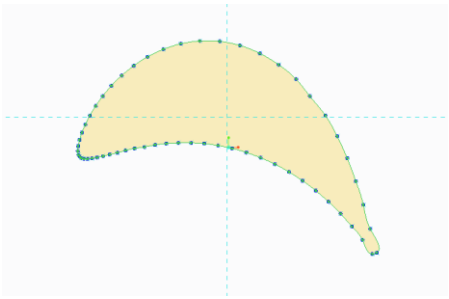
Thrust bearings and journal bearings are a critical part of design. They are hydrodynamic oil bearings or oil-cooled rolling-element bearings. Foil bearings are used in some small machines such as micro turbines and also have strong potential for use in small gas turbines/auxiliary power units.

2. 3D MODELS OF BLADES

2.1 Points

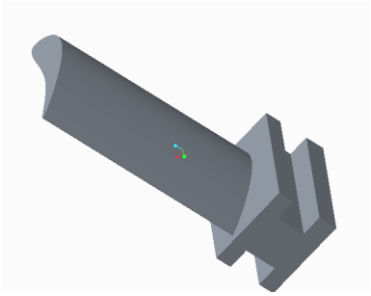


2.2 SKETCHER



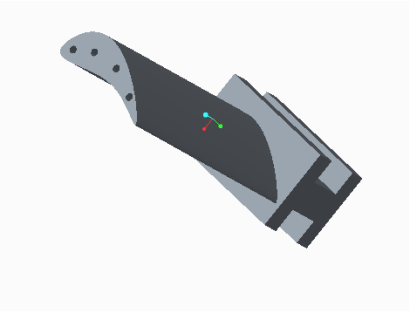
2.3 ORIGINAL MODEL

WITHOUT HOLES

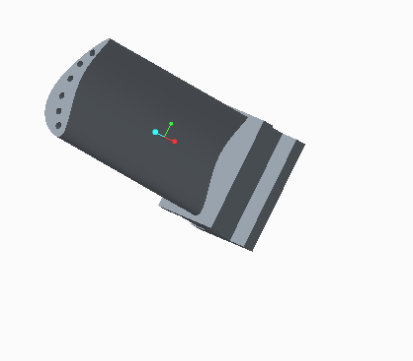


2.4 MODIFIED MODEL

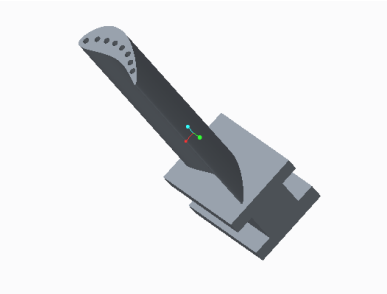
WITH 4 HOLES



2.5 WITH 6 HOLES



2.6 WITH 8 HOLES



3. INTRODUCTION TO FEA

Finite element analysis is a method of solving, usually approximately, certain problems in engineering and science. It is used mainly for problems for which no exact solution, expressible in some mathematical form, is available. As such, it is a numerical rather than an analytical method. Methods of this type are needed because analytical methods cannot cope with the real, complicated problems that are met with in engineering. For example, engineering strength of materials or the mathematical theory of elasticity can be used to calculate analytically the stresses and strains in a bent beam, but neither will be very successful in finding out what is happening in part of a car suspension system during cornering.

One of the first applications of FEA was, indeed, to find the stresses and strains in engineering components under load. FEA, when applied to any realistic model of an engineering component, requires an enormous amount of computation and the development of the method has depended on the availability of suitable digital computers for it to run on. The method is now applied to problems involving a wide range of phenomena, including vibrations, heat conduction, fluid mechanics and electrostatics, and a wide range of material properties, such as linear-elastic (Hookean) behavior and behavior involving deviation from Hooke's law (for example, plasticity or rubber-elasticity).

Many comprehensive general-purpose computer packages are now available that can deal with a wide range of phenomena, together with more specialized packages for particular applications, for example, for the study of dynamic phenomena or large-scale plastic flow. Depending on the type and complexity of the analysis, such packages may run on a microcomputer or, at the other extreme, on a supercomputer. FEA is essentially a piece-wise process. It can be applied to one-dimensional problems, but more usually there is an area or volume within which the solution is required. This is split up into a number of smaller areas or volumes, which are called finite elements. Figure 1 shows a two-dimensional model of a spanner that has been so divided: the process is called discretisation, and the assembly of elements is called a mesh.

4. INTRODUCTION TO ANSYS

Structural Analysis

ANSYS Autodyn is computer simulation tool for simulating the response of materials to short duration severe loadings from impact, high pressure or explosions.

4.1 ANSYS Mechanical

ANSYS Mechanical is a finite element analysis tool for structural analysis, including linear, nonlinear and dynamic studies. This computer simulation product provides finite elements to model behavior, and supports material models and equation solvers for a wide range of mechanical design problems. ANSYS Mechanical also includes thermal analysis and coupled-physics capabilities involving acoustics, piezoelectric, thermal–structural and thermo-electric analysis.

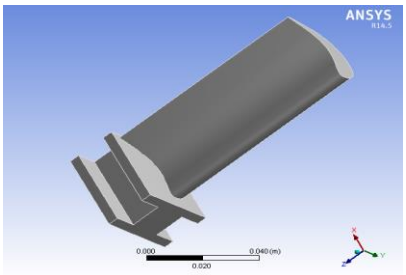
4.2 Fluid Dynamics

ANSYS Fluent, CFD, CFX, FENSAP-ICE and related software are Computational Fluid Dynamics software tools used by engineers for design and analysis. These tools can simulate fluid flows in a virtual environment — for example, the fluid dynamics of ship hulls; gas turbine engines (including the compressors, combustion chamber, turbines and afterburners); aircraft aerodynamics; pumps, fans, HVAC systems, mixing vessels, hydro cyclones, vacuum cleaners, etc.

5. STRUCTURAL ANALYSIS OF GAS TURBINE BLADE

WITHOUT HOLES

TITANIUM ALLOY



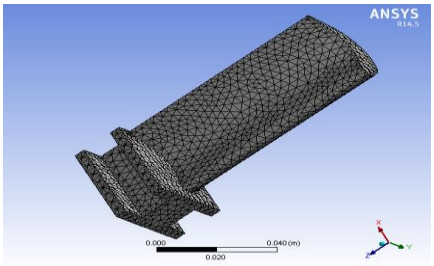
5.1 Material properties of Titanium Alloy

Density : 0.0000134 kg/mm^3

Young's modulus : 125000Mpa

Poisson's ratio : 0.342

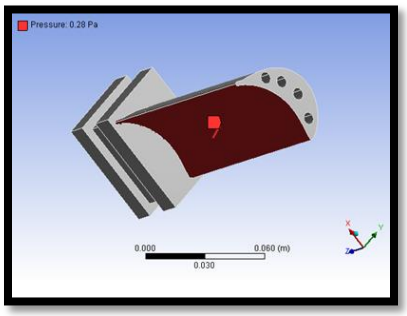
Meshed model



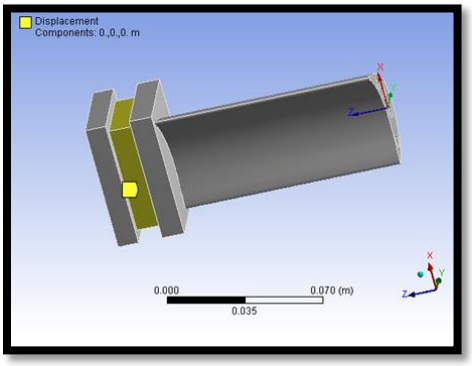
5.2 WITH 4 HOLES

TITANIUM ALLOY

Pressure

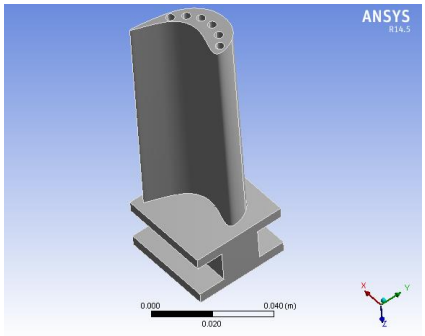


Displacement

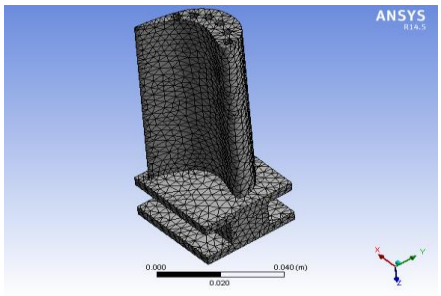


5.3 TITANIUM ALLOY

Save CREO Model as .iges format



Meshed model

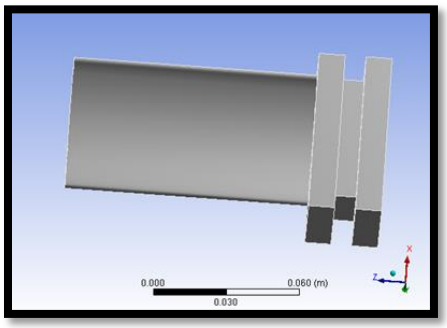


5.4 THERMAL ANALYSIS ON GAS TURBINE BLADE

TITANIUM ALLOY

WITHOUT HOLES

Import model



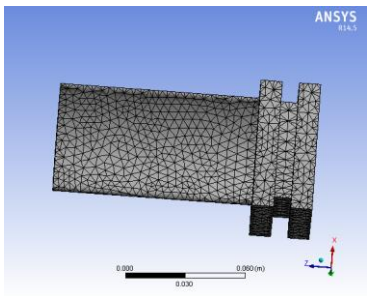
Material Properties of Titanium Alloy

Density : 0.00000484 kg/mm^3

Thermal conductivity : 10.9 w/mk

Specific heat: 670 j/g^^{0}c

Meshed model



6. CONCLUSION

In our project we have designed a turbine blade used in gas turbines and modeled in 3D modeling software Pro/Engineer. Two other models with 4 holes and 6 holes are also modeled.

We have done structural and thermal analysis on all the models of turbine blades using Titanium alloy and Nickel alloy. By observing the analysis results, the analyzed stress values are less than their permissible stress values. So using both the materials is safe. The stress and deformation values are more for Nickel alloy.

By observing the thermal results, thermal flux is more for Nickel alloy than titanium alloy. So using Nickel alloy is better than Titanium alloy. But the main disadvantage is its weight.

By comparing the results for all the models, thermal flux is increasing by increasing number of holes, so heat transfer rate is increased.

So we can conclude that by using Nickel alloy with 6 holes is better.

7. REFERENCES

1. Design and Analysis of Gas Turbine Blade by Theju V, Uday P S , PLV Gopinath Reddy, C.J.Manjunath
2. Heat Transfer Analysis of Gas Turbine Blade Through Cooling Holes by K Hari Brahmaiah , M.Lava Kumar
3. The design and analysis of gas turbine blade by Pedaprolu Venkata Vinod
4. Effect Of Temperature Distribution In 10c4/60c50 Gas Turbine Blade Model Using Finite Element Analysis by V.Veeraragavan
5. Film Cooling on a Gas Turbine Rotor Blade by K. Takeishi, S. Aoki, T. Sato and K. Tsukagoshi
6. Film Cooling of the Gas Turbine Endwall by Discrete-Hole Injection by M. Y. Jabbari, K. C. Marston, E. R. G. Eckert and R. J. Goldstein
7. An advanced impingement/film cooling scheme for gas turbines – numerical study by A. Immarigeon, (Department of Mechanical and Industrial Engineering, Concordia University, Montréal, Canada), I. Hassan, (Department of Mechanical and Industrial Engineering, Concordia University, Montréal, Canada)
8. An experimental investigation on the trailing edge cooling of turbine blades by Zifeng Yang, Hui Hu
9. CFD Simulation on Gas turbine blade and Effect of Hole Shape on leading edge Film Cooling Effectiveness by Shridhar Paregouda, Prof. Dr. T. Nageswara Rao

ANALYSIS OF HEAT TRANSFER RATE BY VARYING COOLING FLUID FOR ENGINE CYLINDER FINS

MOHD IMRAN
Mohdimran9605@gmail.com
M TECH(CAD/CAM)
Malla Reddy College Of Engineering,Mysammaguda
Hyderabad

Abstract: The Engine cylinder is one of the major automobile components, which is subjected to high temperature variations and thermal stresses. In order to cool the cylinder, fins are provided on the cylinder to increase the rate of heat transfer. By doing thermal analysis on the engine cylinder fins, it is helpful to know the heat dissipation inside the cylinder.The principle implemented in this project is to increase the heat dissipation rate by using the invisible working fluid, nothing but air. We know that, by increasing the surface area we can increase the heat dissipation rate, so designing such a large complex engine is very difficult. The main purpose of using these cooling fins is to cool the engine cylinder by air.

Keywords:- Ansys,Catia,Engine Cylinder Fins

1. INTRODUCTION

The internal combustion engine is an engine in which the combustion of a fuel (normally a fossil fuel) occurs with an oxidizer (usually air) in a combustion chamber. In an internal combustion engine, the expansion of the high-temperature and -pressure gases produced by combustion applies direct force to some component of the engine, such as pistons, turbine blades, or a nozzle. This force moves the component over a distance, generating useful mechanical energy.

1.1 NECESSITY OF COOLING SYSTEM IN IC ENGINES

All the heat produced by the combustion of fuel in the engine cylinders is not converted into useful power at the crankshaft. A typical distribution for the fuel energy is given below:

It is seen that the quantity of heat given to the cylinder walls is considerable and if this heat is not removed from the cylinders it would result in the resignation of the charge. In addition, the lubricant would also burn away, thereby causing the seizing of the piston. Excess heating will also damage the cylinder material.

Keeping the above factors in view, it is observed that suitable means must be provided to dissipate the excess heat from the cylinder walls, so as to maintain the temperature below certain limits.

However, cooling beyond optimum limits is not desirable, because it decreases the overall efficiency due to the following reasons:

Thermal efficiency is decreased due to more loss of heat to the cylinder walls.

- The vaporization of fuel is less; this results in fall of combustion efficiency.
- Low temperatures increase the viscosity of lubrication and hence more piston friction is encountered, thus decreasing the mechanical efficiency.

Though more cooling improves the volumetric efficiency, yet the factors mentioned above result in the decrease of overall efficiency.

Thus it may be observed that only sufficient cooling is desirable and any deviation from the optimum limits will result in the deterioration of the engine performance.

2. METHODS OF COOLING

Various methods used for cooling of automobile engines are:

1.Air Cooling 2.Water cooling

2.1 AIR-COOLING

Cars and trucks using direct air cooling (without an intermediate liquid) were built over a long period beginning with the advent of mass produced passenger cars and ending with a small and generally unrecognized technical change. Before World War II, water cooled cars and trucks routinely overheated while climbing mountain roads, creating geysers of boiling cooling water. This was considered normal, and at the time, most noted mountain roads had auto repair shops to minister to overheating engines.

ACS (Auto Club Suisse) maintains historical monuments to that era on the Susten Pass where two radiator refill stations remain (See a picture here). These have instructions on a cast metal plaque and a spherical bottom watering can hanging next to a water spigot. The spherical bottom was intended to keep it from being set down and, therefore, be useless around the house, in spite of which it was stolen, as the picture shows.

During that period, European firms such as Magirus-Deutz built air-cooled diesel trucks, Porsche built air-cooled farm tractors, and Volkswagen became famous with air-cooled passenger cars. In the USA, Franklinbuilt air-cooled engines. The Czechoslovakia based company Tatra is known for their big size air cooled V8 car engines, Tatra engineer Julius Mackerle published a book on it. Air cooled engines are better adapted to extremely cold and hot environmental weather temperatures, you can see air cooled engines starting and running in freezing conditions that stuck water cooled engines and continue working when water cooled ones start producing steam jets.

2.2 LIQUID COOLING

Today, most engines are liquid-cooled.

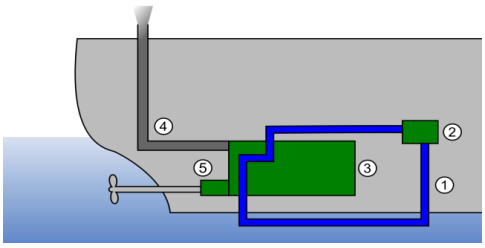


Fig:1 A fully closed IC engine cooling system

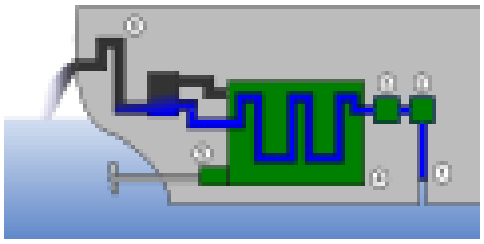


Fig:2 Open IC engine cooling system

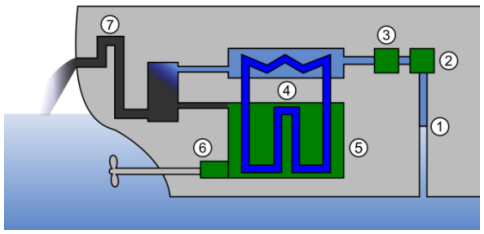


Fig:3 Semi closed IC Engine cooling system

Liquid cooling is also employed in maritime vehicles (vessels, ...). For vessels, the seawater itself is mostly used for cooling. In some cases, chemical coolants are also employed (in closed systems) or they are mixed with seawater cooling.

3. INTRODUCTION OF ANSYS :

ANSYS is general-purpose finite element analysis (FEA) software package. Finite Element Analysis is a numerical method of deconstructing a complex system into very small pieces (of user-designated size) called elements. The software implements equations that govern the behaviour of these elements and solves them

all; creating a comprehensive explanation of how the system acts as a whole. These results then can be presented in tabulated, or graphical forms. This type of analysis is typically used for the design and optimization of a system far too complex to analyze by hand. Systems that may fit into this category are too complex due to their geometry, scale, or governing equations.

ANSYS is the standard FEA teaching tool within the Mechanical Engineering Department at many colleges. ANSYS is also used in Civil and Electrical Engineering, as well as the Physics and Chemistry departments.

ANSYS provides a cost-effective way to explore the performance of products or processes in a virtual environment. This type of product development is termed virtual prototyping.

With virtual prototyping techniques, users can iterate various scenarios to optimize the product long before the manufacturing is started. This enables a reduction in the level of risk, and in the cost of ineffective designs. The multifaceted nature of ANSYS also provides a means to ensure that users are able to see the effect of a design on the whole behavior of the product, be it electromagnetic, thermal, mechanical etc.

3.1 GENERIC STEPS TO SOLVING ANY PROBLEM IN ANSYS :

Like solving any problem analytically, you need to define (1) your solution domain, (2) the physical model, (3) boundary conditions and the physical properties. You then solve the problem and present the results. In numerical methods, the main difference is an extra step called mesh generation. This is the step that divides the complex model into small elements that become solvable in an otherwise too complex situation. Below describes the processes in terminology slightly more attune to the software.

Build Geometry

Construct a two or three dimensional representation of the object to be modeled and tested using the work plane coordinate system within ANSYS.

Define Material Properties

Now that the part exists, define a library of the necessary materials that compose the object (or project) being modeled. This includes thermal and mechanical properties.

Generate Mesh

At this point ANSYS understands the makeup of the part. Now define how the modeled system should be broken down into finite pieces.

Apply Loads

Once the system is fully designed, the last task is to burden the system with constraints, such as physical loadings or boundary conditions.

Obtain Solution

This is actually a step, because ANSYS needs to understand within what state (steady state, transient… etc.) the problem must be solved.

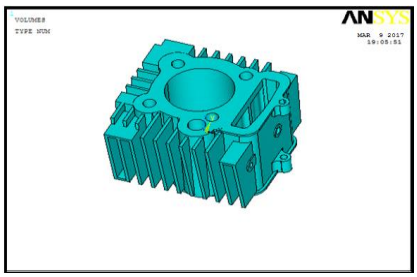
Present the Results

After the solution has been obtained, there are many ways to present ANSYS' results, choose from many options such as tables, graphs, and contour plots.

Thermal analysis of fin body:

Set Units - /units,si,mm,kg,sec,k

File- change Directory-select working folder File-Change job name-Enter job name Preferences-Thermal\preprocessor-Element type-add/edit/delete-Select Add-Solid 20 node 90



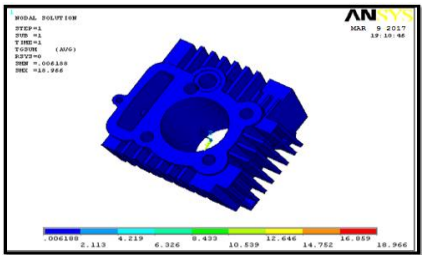
3.2MESHED MODEL OF ALUMINIUM ALLOY 204 3MM THICKNESS

Finite element analysis or FEA representing a real project as a “mesh” a series of small, regularly shaped tetrahedron connected elements, as shown in the above fig.And then setting up and solving huge arrays of simultaneous equations. The finer the mesh, the more accurate the results but more computing power is required.

Nodal temperature of Aluminium Alloy 204 3mm thickness

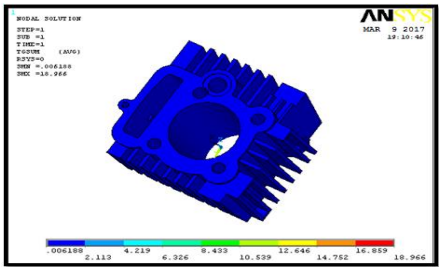
According to the contour plot, the temperature distribution maximum temperature at bore because the operating temperature passing inside of the bore. So we applied the temperature inside of the bore and applied the convection to fins. Then the maximum temperature at bore and its distributed to outer surface of the fins.

General post processer- contour plot- Thermal Gradient-Thermal Gradient Vector Sum



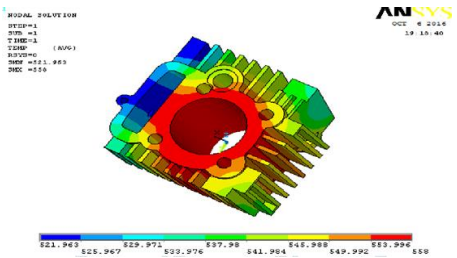
Thermal Gradient of Aluminium Alloy 204 3mm thickness

According to the contour plot, the thermal gradient maximum at bore because the operating temperature passing inside of the bore. So we applied the temperature inside of the bore and applied the convection to fins. Then the minimum gradient at fins. According to the above contour plot, the maximum gradient is 18.966 k/m and minimum gradient is 0.006188 k/m.**General post processer- contour plot- Thermal Flux –Thermal Flux Vector Sum**



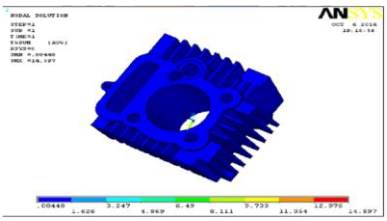
Nodal Temperature of Aluminium Alloy 6061 3mm thickness

According to the contour plot, the temperature distribution maximum temperature at bore because the operating temperature passing inside of the bore. So we applied the temperature inside of the bore and applied the convection to fins. Then the maximum temperature at bore and its distributed to outer surface of the fins.



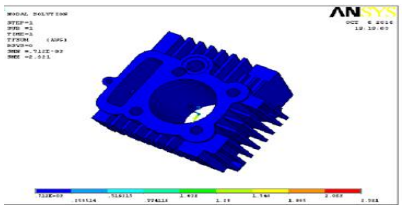
3.3 NODAL TEMPERATURE OF MAGNESIUM ALLOY 3MM THICKNESS

According to the contour plot, the temperature distribution maximum temperature at bore because the operating temperature passing inside of the bore. So we applied the temperature inside of the bore and applied the convection to fins. Then the maximum temperature at bore and its distributed to outer surface of the fins.



3.4THERMAL GRADIENT SUM OF MAGNESIUM ALLOY 3MM THICKNESS

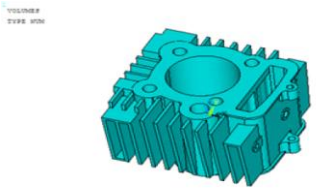
According to the contour plot, the thermal gradient maximum at bore because the operating temperature passing inside of the bore. So we applied the temperature inside of the bore and applied the convection to fins. Then the minimum gradient at fins. According to the above contour plot, the maximum gradient is 14.597 k/m and minimum gradient is 0.00448 k/m.



3.5 THERMAL FLUX SUM OF MAGNESIUM ALLOY 3MM THICKNESS

According to the contour plot, the thermal flux maximum at bore because the operating temperature passing inside of the bore. So we applied the temperature inside of the bore and applied the convection to fins. Then the minimum thermal flux at fins. According to the above contour plot, the maximum thermal flux is 2.321 k/m and minimum thermal flux is 0.712E-03 k/m.

ALUMINUM ALLOY 204 – 2.5mm THICKNESS



MODEL ANALYSIS OF ALUMINUM ALLOY 204 – 2.5mm THICKNESS

MODEL IMPORTED FROM PRO/ENGINEER

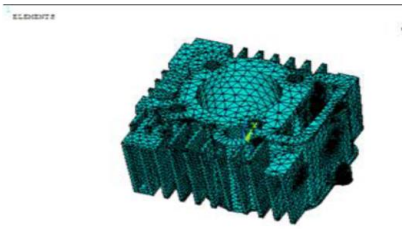
MATERIAL PROPERTIES

Thermal Conductivity – 120 w/mk

Specific Heat – 0.963 J/g ºC

Density – 2.8 g/cc

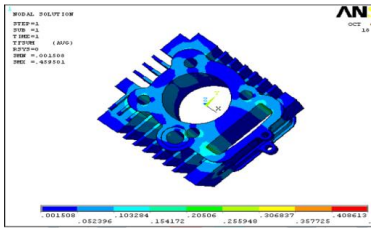
MESHED MODEL



3.6 NODAL TEMPERATURE OF ALUMINIUM ALLOY 204 2.5MM THICKNESS

According to the contour plot, the temperature distribution maximum temperature at bore because the operating temperature passing inside of the bore. So we applied the temperature inside of the bore and applied the convection to fins. Then the maximum temperature at bore and its distributed to outer surface of the fins.

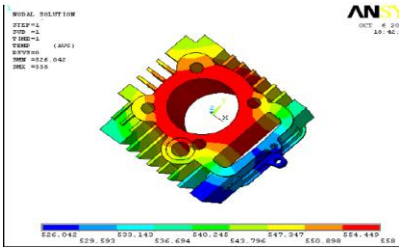
THERMAL FLUX SUM



3.7 THERMAL FLUX SUM SUM OF ALUMINIUM ALLOY 204 2.5MM THICKNESS

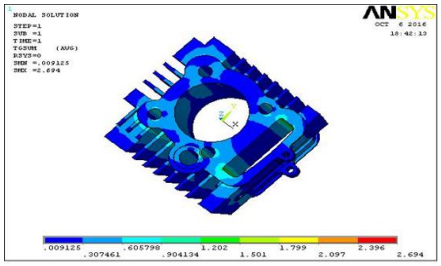
According to the contour plot, the thermal flux maximum at bore because the operating temperature passing inside of the bore. So we applied the temperature inside of the bore and applied the convection to fins. Then the minimum thermal flux at fins.

According to the above contour plot, the maximum thermal flux is 0.459501 k/m and minimum thermal flux is 0.001508 k/m.



3.8 NODAL TEMPERATURE OF ALUMINIUM ALLOY 6061 2.5 MM THICKNESS

According to the contour plot, the temperature distribution maximum temperature at bore because the operating temperature passing inside of the bore. So we applied the temperature inside of the bore and applied the convection to fins. Then the maximum temperature at bore and its distributed to outer surface of the fins.



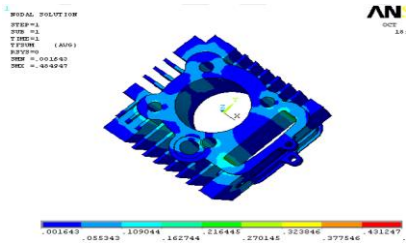
THERMAL GRADIENT SUM OF ALUMINIUM ALLOY 6061 2.5 MM THICKNESS

According to the contour plot, the thermal gradient maximum at bore because the operating temperature passing inside of the bore.

So we applied the temperature inside of the bore and applied the convection to fins. Then the minimum gradient at fins.

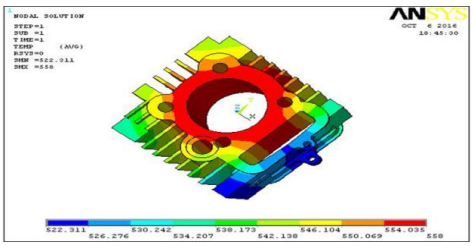
According to the above contour plot, the maximum gradient is 2.694 k/m and minimum gradient is 0.009125 k/m.

THERMAL FLUX SUM



3.9 THERMAL FLUX SUM OF ALUMINIUM ALLOY 6061 2.5 MM THICKNESS

According to the contour plot, the thermal flux maximum at bore because the operating temperature passing inside of the bore. So we applied the temperature inside of the bore and applied the convection to fins. Then the minimum thermal flux at fins. According to the above contour plot, the maximum thermal flux is0.484947 k/m and minimum thermal flux is 0.001643 k/m.



THERMAL FLUX SUM OF MAGNESIUM ALLOY 2.5 MM THICKNESS

According to the contour plot, the thermal flux maximum at bore because the operating temperature passing inside of the bore. So we applied the temperature inside of the bore and applied the convection to fins. Then the minimum thermal flux at fins. According to the above contour plot, the maximum thermal flux is 0.477984 k/m and minimum thermal flux is 0.001604 k/m.

4. CONCLUSION

In this thesis, a cylinder fin body for a 150cc motorcycle is modeled using parametric software Pro/Engineer. The original model is changed by changing the thickness of the fins. The thickness of the original model is 3mm, it has been reduced to 2.5mm. By reducing the thickness of the fins, the overall weight is reduced.

Present used material for fin body is Aluminum Alloy 204. In this thesis, two other materials are considered which have more thermal conductivities than Aluminum Alloy 204. The materials are Aluminum alloy 6061 and Magnesium Alloy. Thermal analysis is done for all the three materials. The material for the original model is changed by taking the consideration of their densities and thermal conductivity.

By observing the thermal analysis results, thermal flux is more for Aluminum alloy 6061 than other two materials and also by reducing the thickness of the fin, the heat transfer rate is increased.

Thermal flux is also calculated theoretically. By observing the results, heat transfer rate is more when the thickness of the fin is 2.5mm.

So we can conclude that using Aluminum alloy 6061 and taking thickness of 2.5mm is better.

5.REFERENCES

[1] Thermal Engineering by I. Shvets, M. Kondak

[2] Thermal Engineering by Rudramoorthy

[3] Thermal Engineering by R.K. Rajput

[4] Thermal Engineering by Sarkar

[5] Online Materials

[6] Gibson, A.H., The Air Cooling of Petrol Engines, Proceedings of the Institute of Automobile Engineers, Vol.XIV (1920), pp.243 – 275.

[7] Biermann, A.E. and Pinkel, B., Heat Transfer from Finned Metal Cylinders in an Air Stream, NACA Report No.488 (1935).

[8] Thornhill, D. and May, A., An Experimental Investigation into the Cooling of Finned Metal Cylinders, in a

[9] Free Air Stream, SAE Paper 1999-01-3307, (1999). (4) Thornhill, D., Graham, A., Cunnigham, G., Troxier, P. and Meyer, R.,

[10] Experimental Investigation into the Free Air-Cooling of Air-Cooled Cylinders, SAE Paper 2003-32-0034, (2003). (5) Pai, B.U., Samaga, B.S. and Mahadevan, K., Some

[11] Experimental Studies of Heat Transfer from Finned Cylinders of Air-Cooled I.C. Engines, 4th National Heat Mass Transfer Conference, (1977), pp.137–144.

REUSING, RECYCLING AND UP-CYCLING OF BIOMASS

CH.SHIVA SAI NARESH
Shivasai36.chittiprolu@gmail.com

CH.BHASKAR
Bhaskar1438480@gmail.com

M.TECH.

Abstract:- **There is a growing interest in the utilisation of biomass for a range of applications. Coupled with this is the appeal of improving the circular economy and as such, there is a focus on reusing, recycling and upcycling of many materials, including biomass. This has been driven by society in terms of demand for more sustainable energy and products, but also by a paradigm shift in attitudes of the population to reduce their personal carbon footprint. Herein we have selected a number of types of biomass (woody, herbaceous, etc.) and surveyed the ways in which they are utilised. We have done this in combination with assessing some kinetic modelling approaches which been reported for the evaluation of different processes for the recycling, reuse and upcycling of biomass.**

Keywords:- General classifications of biomass, Processing of biomass resources, Practical applications, Reusing and recycling, Herbaceous biomass, Pyrolysis and bio-oil production, Other types of biomass sources upcycling

1.Introduction

Worldwide, there are issues around "waste" that are prevalent. Perhaps the most pressing issue is the attitudes and behavioural issues in terms of designating something as "waste". From a chemical, and in turn energy, perspective there is normally an intrinsic (carbon) value in such materials; so that it would be better to deem this as "resources" rather than waste. This way of thinking can be considered part of the circular economy, a model by which as much value as possible is retained via reuse, recycling, repurposing and up-cycling with aspirations towards zero waste. As such, there is interest in evaluating the potential value in the circular economy that was highlighted recently by Zacho [1].

1.1 General classifications of biomass

In general, biomass can be classified as primary, secondary and tertiary, where:

Primary biomass:- It is produced directly via photosynthesis and is taken directly from the land. For instance, herbaceous and woody biomass along with the seeds of oil crops and the residues after the harvesting of forest trees and agricultural crops (corn stover, limbs, bark and wheat straw) are all considered primary biomass source [1].

Secondary biomass:- It results from the processing of the primary bio- mass such as sawdust (physical processing), black liquor (chemical processing) or manure production by animals (biological processing) [1].

Tertiary biomass:- It is post-consumer byproducts such as animal fat, used vegetable oils, construction and demolition debris and packing by products [1].

1.2 Classification of biofuel generations

Biomass, when used as a biofuel, can also be classified to the first, second, third and fourth generation where:

(i) 1st generation biofuels come from edible biomass such as the ethanol production from fermenting sugars. There are major drawbacks of this generation in terms of the food versus fuel debate, low land use efficiency along with the geographical limitations

(ii) 2nd generation biofuels are derived from lignocellulosic biomass, either non-edible of food crop (rice husks) or non-edible whole plant biomass resources (grasses). The major challenge related to this generation is that processing can be expensive and a technological breakthrough is still required to make them more feasible.

(iii) 3rd generation biofuels are produced using non-arable land such as algae and typically need a high capital cost to operate

(iv) 4th generation biofuels can be made using non-arable land and do not require the destruction of biomass to be converted into fuel such as the photobiological solar fuels [1].

1.3 Processing of biomass resources

Typically biomass resources can be processed in a number of ways such as:

(i) Combustion, which is the burning of biomass in the air to produce hot gases, ash and consequently converting the stored chemical energy into heat which can be then converted to kinetic energy through heating water to produce vapour which is used for gas engines.

(ii) Gasification, which is the partial oxidation of biomass (for ex- ample rice husk) at high temperatures to convert the biomass to a combustible gas mixture which can be burnt directly or used as a fuel for gas engines, gas turbines or can be used for the production of chemicals in the case of the medium energy gases.

(iii) Pyrolysis, which is the thermal degradation of the biomass in an oxygen-free environment which results a char (solid product), condensable vapour (bio-oil) along with the gaseous products

(iv) Liquefaction, which is the conversion of biomass into stable liquid hydrocarbons at low temperatures and high hydrogen pressures [1].

2. Introductory piece to reusing and recycling approaches

2.1 Practical applications

Firstly, using biomass for combustion in the energy sector depends upon the quality, chemical and physical parameters of the feedstock or feedstocks in question. Mancini et al. developed a combined technique of near-infrared spectroscopy with multivariate analysis to predict the gross calorific value (GCV) and the ash content in the bio- mass feedstock, with a standard error of 0.23 $kJ \cdot g^{-1}$ and 0.4%, respectively. This approach could be used on-line or off-line for providing an indication of the quality of the feedstock and its suitability for the industrial application. Unlike coal, biomass is characterised by its larger volatile content while also being a porous material i.e. biomass allows the ingress of bacteria and O_2, thus, increasing the risk of self-heating and ignition in storage and transportation. Therefore, monitoring the temperature, volatile organic compounds (VOC), carbon monoxide (CO) and controlling the moisture is crucial to avoid the ignition [2].

2.2 Introductory piece on kinetic modelling

Understanding the thermal kinetic behavior of the thermal de- composition of biomass is crucial in order to identify the physicochemical combustion characteristics that hinder some of the energy generation applications. Thus, determining the activation energy (E_a) and the pre-exponential factor using either model-fitting or model-free (isoconversional) approaches, which are the two common methods for studying the Differential scanning calorimetry/Thermogravimetric analysis (DSC/TGA) solid-state kinetic data, is important. The model- fitting method is based on the best fit of the different models with the experimental thermogravimetric data using a single TGA/DSC curve, which might actually fit with more than one model. Conversely, the model-free isoconversional method is a simple method that eliminates the error related to the model fitting. In the isoconversional method, different heating rate curves are required to calculate the kinetic parameters as a function of the extent of conversion (α) i.e. the E_a is calculated for a serious of different conversion points. It is well known that the thermal analysis mechanism is complicated as it tends to take

place in multiple steps with different reaction rates and, as such, an isoconversional method is the most suitable method and, consequently, the most commonly used in this case. As biomass is pre- dominantly composed of three different components (cellulose, hemi- cellulose and lignin), each behaves differently and separately during the biomass thermal decomposition. Thus, this complex behaviour of bio- mass decomposition should be considered in kinetic modeling [2].

3. Reusing and recycling

3.1. Woody biomass

There are significant opportunities to recycle and reuse wood. Many items, such as furniture, can be sold second-hand and used again; and there has been human behavioural/attitude changes with regards to secondhand consumption/use which had been typical during the19th Century, but became stigmatized in the 20th Century [3].

More recently there has been a marked increase in secondhand use, which may have been financially driven by the impact on consumers from the 2008 economic crisis. This was likely not the only driver in such a paradigm shift. Consumers are typically much more informed now than previous generations and are aware of environmental issues including land use concerns such as deforestation for agriculture, the eat versus heat debate and landfill capacity. As such, consumers are very conscious of their carbon footprint and, consequently, this is also being factored into their decisions. For example, it has been reported that the reuse/recycle pathways can help reduce GHG emissions, thereby reducing environmental impact

Generally, with the reuse and recycling of woody biomass there is little, if any, chemical conversion and so there are not many examples of kinetic studies or modelling for such processes. Given that ability of biomass to be converted to solid, liquid and gas fuels, biomass is a suitable energy resource for heat, power and transport.

While the utilisation of liquid and gaseous biofuels typically requires some chemical transformation/valorisation of the biomass, this is not necessarily the same for solid fuels [3].

In order to use unprocessed biomass as a solid fuel, it is normally blended with other fuels, such as coal, in order to compensate for the low calorific value and high moisture content inherent in the biomass source. Woody biomass, can, in theory, be combusted without much processing and as such, the application of woody biomass as a fuel can still be considered a low-grade use [3].

3.2 Herbaceous biomass

Herbaceous biomass usually contains components such as seeds, cones, leaves and stems that require further size reduction as a post-harvest handling procedure to increase the bulk density of the biomass, thus facilitating its storage and transportation alongside increasing the surface area, hence, increasing the chemical process reaction rate. As discussed previously, size reduction methods such as the grinding process require energy. Naimiet al, studied the required energy input for five herbaceous biomass (miscanthus, wheat straw, switch grass, corn stover and canolastraw) and found out that it was in the range of 22–35 kWh·t−1 at a lab
scale while using an industrial grinder showed a required energy input of 59.4 $kWh \cdot t^{-1}$ [3].

Herbaceous biomass such as miscanthus is typically characterized based on its silica and inorganic contents. These components cause various problems during the thermal decomposition via valorization and melting of the low potassium silicates compounds. In the case of woody biomass this problem is less noticeable due to the low alkali and Si content in the biomass. Thus, herbaceous biomass is usually mixed with other woody fuel or coal to harness the energy inherent in these types of biomass. Furthermore, the high ash content in the herbaceous biomass is problematic as it catalyses the thermochemical reactions and affects the yield of the pyrolysis products and their subsequent composition as well. Forbes et al. studied the physicochemical characteristics of eight different biomass fuels such aspine, spruce, brash, Wr (riddled willow), Ws (open air dried chipped willow), miscanthus and commercially available wood pellets. They reported that the standard wood pellet fuel showed the best combustion performance while miscanthus showed the largest clinker formation due to the highest ash content [3].

Biomass-coal co-firing has been shown to diminish the dangerous greenhouse gas emissions and particulates derived from the coal-fired power

plants. However, due to the low bulk, low energy density and wet nature of biomass along with the seasonal availability, it makes it a more challenging process. Co-HTC(miscanthus-coal) showed increased higher heating values (HHV)compared with miscanthus (27.3 and 16.81 MJ·kg–1, respectively).While it decreased regarding the pure coal, due to the low mass density with HHV of 28.10 $MJ \cdot kg^{-1}$ [4].

3.3 Other types of biomass

Species of seaweed are conventionally characterised by their physical colour. The contents of organic based constituents such as carbohydrates, fats and proteins vary considerably in different types and species of seaweeds. For instance, brown seaweed is very rich in carbohydrates with its protein content is relatively low, whereas approximately 33 wt% of red seaweed is protein. This feature of brown seaweeds may be useful in carbonisation based processes since carbohydrates can produce volatile gaseous compounds, as well as some char in the pyrolysis/carbonisation processes. Phaeophyta or "brown seaweed" as it is more commonly known, is the most common type of seaweed. Mannitol, fucoidan, laminarin and alginic acid are the main building blocks of this particular type of seaweed.

Conventionally, seaweed is classed as unsuitable for thermochemical conversions such as combustion and gasification unless pre-treatments are carried out or the application is carried out in conjunction with a co-existing feedstock. It is the inorganic matter that is contained within the seaweed that gives rise to some problems. Typically, this will lead to a higher variation in ash-forming elements, ash content and high levels of salts (e.g. sodium chloride) compared to other fuel sources available such as coal or diesel [5].

The combination of high Na and Cl concentrations promote the risk of alkali chloride-related operational problems. There are several studies reporting on the complex ash composition that arises from seaweed-based feedstocks. For example, compared to land-based biomass, the composition of seaweeds are heterogeneous with high concentrations of ash-forming elements, higher ratios of Na/K and Mg/Ca and higher Cl contents. These higher concentrations can be explained by the region and environment in which the biomass grows (i.e. sea water). Characterisation for this type of biomass and its ash content is extremely important because numerous problems (fouling, deposit formation or slagging) can arise when using as a fuel application. These problems will likely cause a financial and time burden in the form of potential shutdowns of apparatus or there being substantial periodical maintenance required. In most cases, seaweeds require a pre-treatment step prior to utilisation, and washing/leaching with water or weak acids are the most common practice to remove mineral matter and halogens

4. Up-cycling approach

4.1. Herbaceous biomass

(i) Pyrolysis and bio-oil production

Recently, the production of bio-oil from biomass has gained great attention as an alternative to fossil fuels. Usually bio-oil, which is typically dark-brown, is composed of different chemicals such as esters ,acids, alcohols, ketones, aldehydes, phenols and lignin-derived oligomers. Bio-oil is thus characterised by its high acidity (pH 2–3), high moisture and viscosity [212]. Bio-oil typically has low heating value compared with fossil fuel with HHV of approximately 20 and40 MJ·kg–1, respectively. In general, the pyrolysis of hemicellulose and cellulose occurred faster than that of lignin with weight loss in the range of 220–315 °C, 315–400°C and 160–900 °C, respectively [5] .

The pyrolysis of cellulose is endothermic while pyrolysis of hemicellulose and lignin is exothermic . Previous studies reported four stages during the pyrolysis; water desorption followed by three decomposition stages for cellulose, hemicellulose and lignin. The product distribution depends upon the feedstock chemical and biochemical composition, the biomass taxonomy and the secondary interaction between liquid and char [5].

In general, the highest yield of liquid (mostly an hydro sugars) is produced by the decomposition of cellulose[6]
along with the lowest char yield. While hemicellulose decomposes to the highest yield of gas and liquid (mainly water, ketones and phenols), lignin produces the phenols in the liquid form along with the highest yield in char.

Woody biomass decomposes slower than that of the herbaceous biomass with less evolution of volatile matters due to the larger contents of cellulose and hemicellulose within the herbaceous biomass. The produced biochar is characterised with higher calorific value than the raw biomass with values in the range of 25–26 MJ·kg−1 which is due to the increase in the fixed carbon content. The bio-oil composition and yield are unfavourably affected by the amount of the ash content in the biomass feedstock, therefore washing it with water, surfactant or either acid improves both the yield and the composition of the bio-oil. For instance, washing the biomass with water showed an effective way to reduce the inorganic materials such as (K, P and Cl) [5] and it is an ideal pre-treatment choice for woody biomass (low ash content). Again, washing with a surfactant such as Triton X-100 can effectively reduce the ash content in comparison to the untreated biomass. Washing the raw biomass with HCl reduced the metal content and consequently increased the volatile materials during the pyrolysis which in turn increased the bio-oil yield and also reduced the hemicellulose present in the polymer structure of the biomass [5].

Banks et al. reported that using stronger acid could lead to hydrolyse both cellulose and hemicellulose completely that eventually decreased the bio-oil yield [5].

4.2 Woody biomass

Woody biomass can be upcycled/valorised via torrefaction, gasification, hydrothermal carbonisation, biological treatments, pyrolysis, anaerobic digestion, fermentation and transesterification. Many of these processes have been described herein previously for recycling/
reusing and/or applied to other types of biomass.

Primarily in this section, the focus will be on the bio-refinery and the extraction of platform and value-added chemicals from woody biomass. Woody biomass can be fed into anaerobic digesters for the conversion of solid/liquid heterogeneous biomass mixtures to biogas, though would typically require a higher residence time compared to some other types of biomass. Pyrolysis and hydrothermal liquefaction result in bio-oil which can then be upgraded to the value added chemicals or fuels. Pyrolysis is conducted between 450 and 500 °C for short residence times at atmospheric conditions and requires drying whereas the hydrothermal liquefaction is at lower temperatures (300–400 °C) at higher pressure (up to 25 MPa) and for longer times (up to 1 h) with the advantage of no drying step being required.

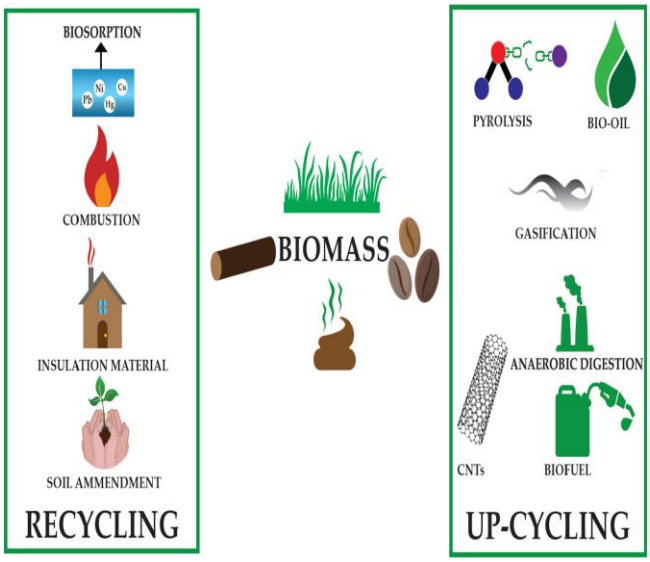
4.3 Other types of biomass sources upcycling

Ross et al. have noted that the char yield from the slow pyrolysis/carbonisation of macro-algae should not be neglected and should be utilised. The abundance of alkali metals in the biomass directly affects the char yield as it alters the pyrolysis mechanism as reported by Ross et al. themselves elsewhere [6].

In order to further predict the yield of pyrolysis products from slow pyrolysis and compare with that of fixed-bed reactor experiments, a mechanistic-based model was designed. This was an adaption of a model originally developed by Van de Velden et al.]. The model
was found to give good accuracy at predicting the pyrolysis products over the temperature range of 400–600 °C producing a mean squared error of 0.49, 16.34 and 13.37 for gas, bio-oil and char, respectively [6].

The overall bio-oil yield from the slow pyrolysis process of cassava peel was found to be 38.7–51.2 wt%, with the optimum yield of 51.2 wt%, corresponding to a temperature of 525 °C. Above this temperature, the yield was found to decrease and this was attributed to secondary thermal cracking. The resultant bio-oil properties satisfied all the ASTM D7544 standard requirements which suggest that bio-oil coming

from the pyrolysis of cassava peel could be used as a resultant bio-fuel.



5. Prospective overview and conclusion

A number of different types of biomass have been considered, while various processes have been discussed for their reuse, recycling an upcycling as seen in Fig. 1. Additionally, the prospects of boosting the circular economy of biomass have been considered. There are significant literature contributions, of which the references of the current work are not an exhausted list, and it is clear that there is a strong desire for maximum utilisation of biomass. By also considering the kinetic modelling of some processes it is hoped to demonstrate that such work can and does play a vital role in the understanding of biomass processing technology. It has also demonstrated that there is still potential to improve on both the actual technologies as well as the kinetic models. Finally, while it is clear that biomass is still required for some lower grade uses, it is preferable to extract the maximum value from the biomass sources if we are truly to achieve a circular economy. It is clear, however, that there is still some progress which is required in this field due to the energy-intensive nature of some of the processes.

6. References

[1] K.O. Zacho, M. Mosgaard, H. Riisgaard, Capturing uncaptured values — a Danish case study on municipal preparation for reuse and recycling of waste, Resource. Conserv. Recycle. 136 (2018) 297–305.

[2] P. Schroeder, P. Dewick, S. Kusi-Sarpong, J.S. Hofstetter, Circular economy and power relations in global value chains: Tensions and trade-offs for lower income countries, Resource. Conservation. Recycle. 136 (2018) 77–78.

[3] Queen's University Belfast gets £1m to research plastic waste, https://www.bbc. co.uk/news/uk-northern-ireland-46648020 , Accessed date: 30 January 2019.

[4] J.M. Encinar, J.F. González, Pyrolysis of synthetic polymers and plastic wastes. Kinetic study, Fuel Process. Technol. 89 (2008) 678–686.

[5] A.I. Osman, J.K. Abu-Dahrieh, F. Laffir, T. Curtin, J.M. Thompson, D.W. Rooney, A bimetallic catalyst on a dual component support for low temperature total methane oxidation, Appl. Catal. B Environ. 187 (2016) 408–418.

[6] A.I. Osman, J. Meudal, F. Laffir, J. Thompson, D. Rooney, Enhanced catalytic activity of Ni on η-Al 2 O 3 and ZSM-5 on addition of ceria zirconia for the partial oxidation of methane, Appl. Catal. B Environ. 212 (2017) 68–79.

DESIGN AND FINITE ELEMENT ANALYSIS OF A PISTON OF INTERNAL COMBUTION ENGINE

Saiteja Singoju, Rama Krishna Pandiboina
Department of Mechanical Engineering,
Malla Reddy College of Engineering,
Jawaharlal Nehru Technological University, Hyderabad.
saitejasingoju@gmail.com
ramakrishnapandiboina@gmail.com

***Abstract*— This journal describes the stress distribution of the piston in four stroke engines by using FEM. Our main objectives is to Study and analyze the thermal stress and maximum or minimum principal stresses, Vanishes stresses distribution on engine piston during combustion process. The journal describes the optimization techniques with using finite element analysis method (FEM) to predict the higher stress and critical region on the component. The stress concentration on the piston head, piston skirt and piston sleeve are reduced by the optimization, using computer aided design (AUTO CAD), Pro- ENGINEER/ CREO software the structural model of a piston will be developed. Furthermore, the FEM analysis is done using Computer Aided Simulation software.**

Keywords— FEA, Pro-E, ANSYS, Piston crown, Piston skirt, stress concentration, Thermal analysis.

I. INTRODUCTION

A piston is a cylindrical engine component that Reciprocates in the cylinder bore by forces produced during the combustion process. The combustion chamber is made gas-tight by the piston rings. In an engine, its transfer force from expanding gas in the cylinder to the crankshaft via connecting rod. As a main part in an engine, piston endures the cyclic gas pressure and the inertial forces at work, and this real working condition may cause the fatigue damage of piston, such as piston skirt wear, piston head or crown cracks and so on. The investigations denote that the greatest stress appears on the upper end of the piston and stress concentration is one of the mainly reason for fatigue failure. On the other hand piston over heating-seizure can only occur when something burns or scrapes away the oil film that exists between the piston and the cylinder wall. Understanding this, it's not hard to visually why oils with exceptionally high film strengths are very desirable. Good quality oils will offer provide a film that stands up to the most intense heat and the pressure loads of a modern high output engine. Thermal analysis is a branch of materials science where the properties of materials are studied as they change with temperature. Finite element method (FEM) are commonly used for thermal Analysis. Due to the complicated working environment for the piston; on one hand, the finite element method (FEM) for the piston became more difficult, on the other hand, though there have many methods which are put forward to apply optimal design, the optimal parameters is not easy to determine. In this, the piston is used in low idle and rated speed gas engine. In order to enhance the engine performance.

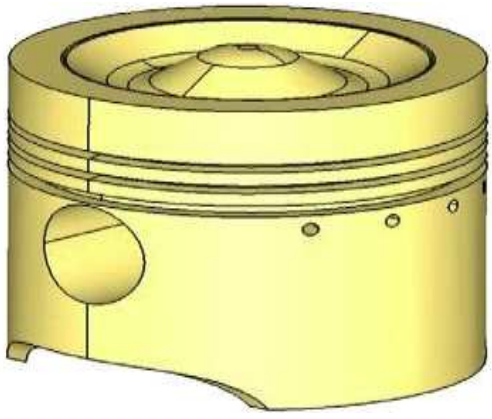


Fig.1 Piston of IC engine

It Is necessary for the piston to be optimized. The mathematical model of optimization is established firstly, and then the FEA is carried out by using ANSYS software. Based on the analysis of optimal result, the stress concentrates on the Upper end of piston has to be evaluated, which provides a better reference for redesign of the piston.

A. Properties of a Piston Material :

Aluminum silicon alloys are widely used in the production of pistons because of their

- High Strength,
- Low Density,
- High Thermal Conductivity,
- Good Cast Ability,
- Workability,
- Good Machinability ,
- High Temperature Resistance.

B. *Aspects of a IC Engine Piston:*

The Piston of an IC engine should comprises of the following characteristics:

- High Strength to resist gas pressure.
- Should have minimum weight.
- Should able to reciprocate with minimum noise.
- Should seal the gas from top and oil from the bottom.
- Should disperse the heat generated during combustion.
- Should have good resistance to distortion under high loads and high temperature.

In this study the piston material we choose is AlSi12CuMgNi cast alloy with eutectic microstructure. As all the engine components around the combustion chamber experience significantly high temperatures and temperature gradients, the temperature dependent property materials has to be used.
According to the thermal analysis results maximum piston temperature reaches approximately 374°C, therefore the cyclic behavior of material is considered at 20, 150, 250 and 350 °C.

C. Engine Specifications:

The engine we are considering in this article is a four-cylinder four -stroke air cooled type Bajaj Kawasaki diesel engine.
The engine specifications are mentioned below:

PARAMETERS	VALUES
Engine Type	Four stroke, diesel engine
Induction	TCIC
Number of cylinders	4 cylinder
Bore	74 mm
Stroke	70 mm
Length of connecting rod	97.6 mm
Displacement volume	99.27 cm^3
Compression ratio	16
Maximum power	21.6 KW at 7000 rpm
Maximum Torque	86 Nm at 3500 rpm
Number of revolutions/cycle	2

Table 1. Engine specifications

II. CAE APPROACH

Computer-Aided-Engineering (CAE) is the computer software to aid in engineering in analyzing tasks. In this standard approach conception ideas are converted into sketches or engineering drawing. With the assistance of this drawings, the prototypes i.e. product which appearance as that of final product are created. It is launched in the market after testing of prototype which gives acceptable results. The thing is, product is launched after doing several practical testing and many trial and error procedures which consumes more time and cost too [1]. Figure 2 depicts the flow process adopted for typical design approach.

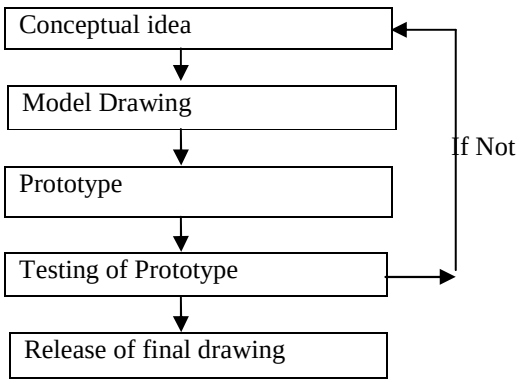


Fig.2 Flowchart of a Conventional Approach to Finite element analysis

FEA method is the mathematical idealization of authentic system. It is a computer predicated method that breaks geometry into element and link a series of equation to every, which are then solved simultaneously to evaluate the outward behavior of the complete system. It is utilizable for perplexed geometry, loading, and material properties where exact analytical solution are difficult to obtain. Most often utilized for structural, thermal, fluid analysis and simulation.

III. METHODOLOGY

During the working cycle operation, the piston is exposed to the high gas pressure and high temperature because of combustion. Simultaneously the piston is fortified by a minute terminus of the connecting rod with the Gudgeon pin. Therefore the methodology for analyzing the piston is considered as the gas pressure given 180 bar is applied uniformly over the crown (top surface/face of the piston) and all degrees of liberation for nodes at upper moiety of piston pin bossed in that piston pin is going to fine- tune. By Considering the fit between piston pin and piston is clearance fit. Only the upper moiety of piston pin boss is considered to be fine-tuning during the analysis.

A. Material properties of piston:

- Material of Piston: - Cast aluminum alloy 201.0
- Young's Modulus [E] – 71 GPa
- Poisson's ratio [µ] – 0.33
- Tensile strength – 485 MPa
- Yield strength – 435 MPa
- Shear strength –290 MPa
- Elongation – 7 %

B. The geometrical representation of an IC Engine piston used for FEA.:

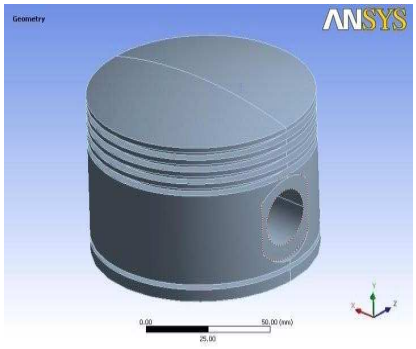


Fig. 4 Basic Geometry of a IC engine piston

A geometrical model of piston is prepared by modeling software's like PRO-e/ CREO, CATIA V5 and be modeled and analysed in the analysis software ANSYS.

C. Finite element model:

The element selected for meshing of the piston model is SOLID187 tetrahedron type of element which is a higher order tetrahedral element. The mesh count for the selected model is 71,910 of nodes and 41,587 number of elements. The figure below is the meshed model of the piston.

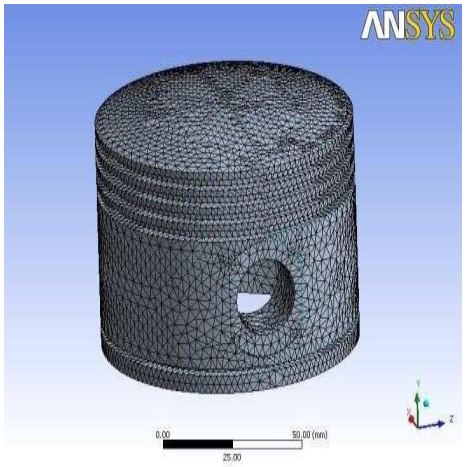


Fig.5 Meshed model of piston

D. Loading & boundary conditions:

The loading and boundary conditions considered for the analysis are showed in the below figure. The uniform pressure of 18 MPa is applied on crown of the piston (top red color) and the model is constrained on upper moiety of piston pin aperture as shown by violet color.

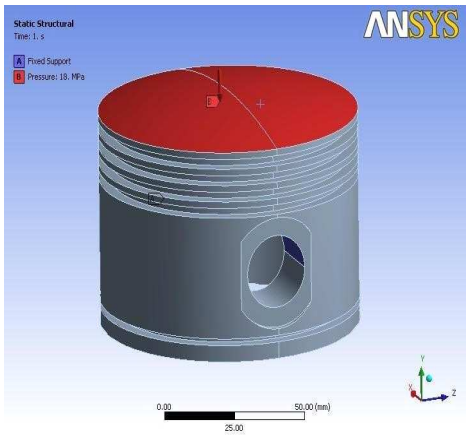


Fig.6 Loading and boundary conditions on piston

IV. RESULTS AND DISCUSSION

A. *Total deflection:*

The maximum deflection in the piston geometry due to the application of gas pressure observed at the central portion of the piston crown is is 0.29669 mm.

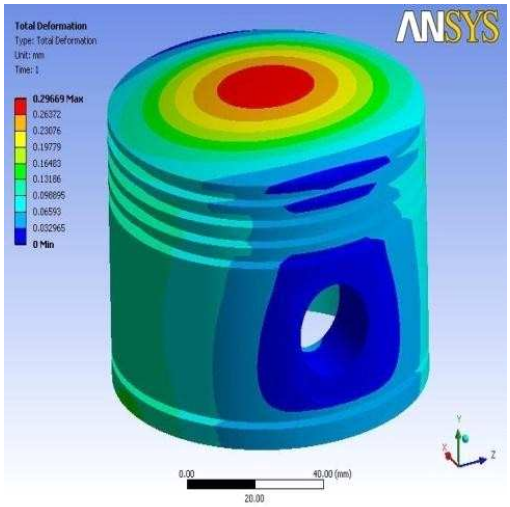


Fig.7 Total deflections on the piston head

B. *Maximum principal stress:*

The distribution of localized and observed at inner side of piston pin boss in the figure below. The overall maximum stresses in the piston body at the inner side of piston crown and piston boss is resulted as 231.25N/mm2.

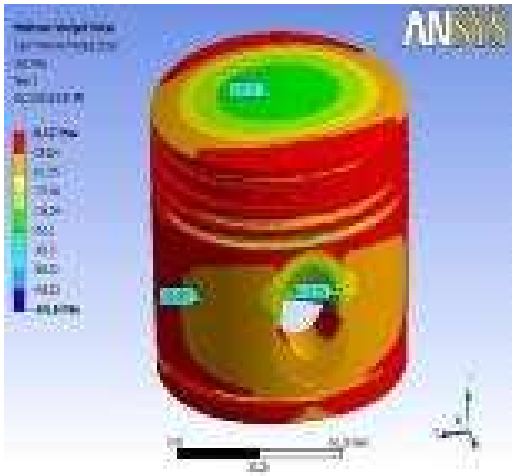


Fig.8 Maximum principal stress on piston

C. *Minimum principal stress:*

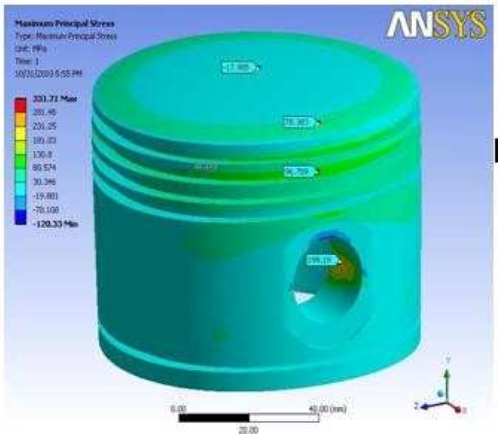


Fig.9 minimum principal stress on piston

The above figure shows the distribution of the minimum principle stresses induced within the piston body. The most maximum values of equivalent stresses are goes up to - 376.74 N/mm^2, which are highly localized and observed at inner side of piston crown & skirt junction. The overall maximum stresses in the piston body is - 250.5 N/mm2 at the top of piston crown.

D. *Von mises stresses:*

The above Figures 10.1 and 10.2 show the distribution of Von mises stresses induced within the piston body. The utmost maximum values of equivalent stresses are goes up to 200.97N/mm2.

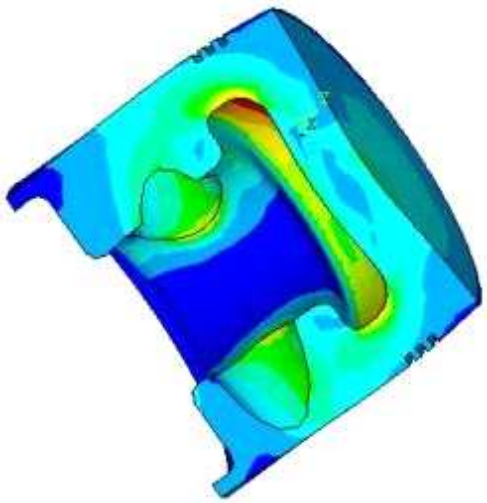


Fig. 10.1 Von mises stress

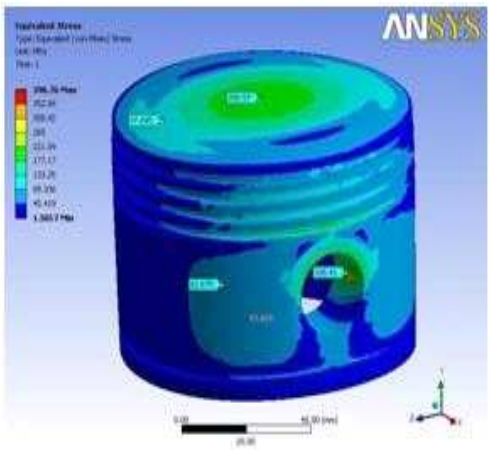


Fig. 10.2 Von mises stress

E. *Stress distribution on the piston body:*

The critical area is observed on the piston head and piston pin hole region. Figure 11 shows stress concentration at various point in piston.

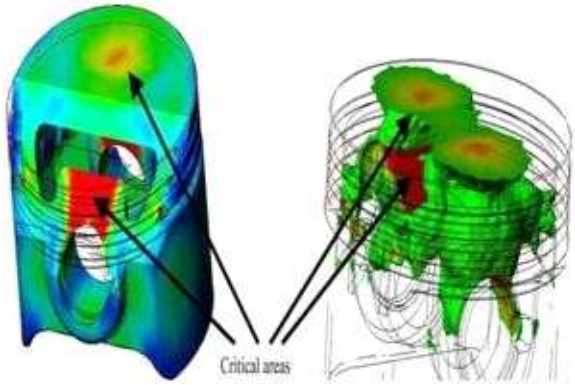


Fig.11 Typical stress distribution on an engine piston

V. CONCLUSION

The Piston skirt may appear deformation during the cyclic operation, which usually results in the crack on the upper end of piston head. Due to the deformation, the greatest stress concentration is caused on the crown, it may leads to the failure of the piston when the stiffness of the piston is not enough, and the crack generally appeared at the point A which may gradually expand and even cause splitting aacross the piston vertically. The stress distribution on the piston mainly depends on the deformation of piston. Therefore, in order to reduce the stress concentration, the piston crown should have enough stiffness.

Also from analysis various results are obtained like The maximum deflection occurred about 0.29669mm due to the application of 180bar gas pressure on crown of piston, 231.25N/mm^2 of maximum principal stress is ascertained, -250.5 N/mm2 of minimum principal stress is ascertained. Also von mises stress of 200.97N/mm2 is observed.

REFERENCES

"Computer Aided Engineering in Compression Molding" (Antoine Rios, Bruce Davis and Paul Grumman, the Madison Group: Polymer Processing Research Corporation 505 S. Ronsard. Madison, WI 53719)

C.H. Li. Piston thermal deformation and friction Considerations, SAE Paper 820086, 1982.

Properties and Selection:High- performance Alloys, ASM Handbook, vol. 1, ASM International, 1990.

Handbook of IC Engines, SAE International.

Martin, Gary G. Failure of stationary pump engine piston, J Failure Anal Prevention 2004;4(1):37–9.

"Fatigue on engine pistons–A compendium of case Studies", F.S.Silva *, 31 March 2005.

FE Analysis of a Vibrating Rigid Circular Piston in Water" by K. Jagadeesh and M. S. Vijaya2SASTECH Vol. VI, NO.1, in the year April 2007.

F.S. Silva, Fatigue on engine pistons – A compendium of case studies, 13 (2006) 480–492.

"Finite Element And MBD Analysis Of Piston To Predicts The Engine Noise" S.N Kurbet1,Vinay V Kuppast1 and Vijay kumar N Chalwa Int. J. Mech. Eng. & Rob. Res. 2013.

Proceedings of 5th International Conference on Future Technologies in Mechanical Engineering - ISBN: 978-81-939386-0-7

DESIGN AND ANALYSIS OF A FOUR WHEELER CRANK SHAFT BY DIFFERENT ALUMINUM ALLOYS

DHEERAVATH SHEKHAR
M.Tech(Thermal Engineering)
Mechanical Engineering Department,
Malla Reddy College Of Engineering,Hyderabad.

***Abstract*:** A crankshaft is used to convert reciprocating motion of the piston into rotary motion or vice versa. The crankshaft consists of the shaft parts which revolve in the main bearings, the crankpins to the big ends of the connecting rod are connected, the crank arms or webs which connects the crankpins and the shaft parts. The crankshafts are subjected to shock and fatigue loads. Thus material of the crankshaft should be tough and fatigue resistant. The common materials used for crankshaft are Carbon Steel or Nickel-Chrome-Moly alloy steel or Nickel-Chrome or special cast iron. The aim of the project is to design and manufacturing a crankshaft for a four cylinder IC engine by using theoretical calculations in design area for Aluminum alloys6061 and commercial grade. A 2D drawing is drafted for crankshaft from the calculations and a 3D model is created in the 3D modeling software CREO

Introduction to crankshaft:

The crankshaft, sometimes casually abbreviated to crank, is the part of an engine which translates reciprocating linear piston motion into rotation. To convert the reciprocating motion into rotation, the crankshaft has "crank throws" or "crankpins", additional bearing surfaces whose axis is offset from that of the crank, to which the "big ends" of the connecting rods from each cylinder attach. It typically connects to a flywheel, to reduce the pulsation characteristic of the four-stroke cycle, and sometimes a torsional or vibrational damper at the opposite end, to reduce the torsion vibrations often caused along the length of the crankshaft by the cylinders farthest from the output end acting on the torsional elasticity of the metal.

1) CRANKSHAFT MANUFACTURING PROCESSES:

Many high performance crankshafts are formed by the forging process, in which a billet of suitable size is heated to the appropriate forging temperature, typically in the range of 1950 - 2250°F, and then successively pounded or pressed into the desired shape by squeezing the billet between pairs of dies under very high pressure. These die sets have the concave negative form of the desired external shape. Complex shapes and / or extreme deformations often require more than one set of dies to accomplish the shaping.

Originally, two-plane V8 cranks were forged in a single plane, then the number two and four main journals were reheated and twisted 90° to move crankpins number two and three into a perpendicular plane. Later developments in forging technology allowed the forging of a 2-plane "non-twist" crank directly (**Figure 1**).

Figure:1
Two-Plane V8 Crankshaft Raw Forging

Crankshafts at the upper end of the motorsport spectrum are manufactured from billet. Billet crankshafts are fully machined from a round bar ("billet") of the selected material (**Figure 2**). This method of manufacture provides extreme flexibility of design and allows rapid alterations to a design in search of optimal performance characteristics. In addition to the fully-machined surfaces, the billet process makes it much easier to locate the counterweights and journal webs exactly where the designer wants them to be. This process involves demanding machining operations, especially with regard to counterweight shaping and undercutting, rifle-drilling main and rod journals, and drilling lubrication passages. The availability of multi-axis, high-speed, high precision CNC machining equipment has made the carved-from-billet method quite cost-effective, and, together with exacting 3D-CAD and FEA design methodologies, has enabled the manufacture of extremely precise crankshafts which often require very little in the way of subsequent massaging for balance purposes.

Figure:2
Billet Crankshaft Machining (Courtesy of Bryant Racing)

There is an old argument that a forged crank is superior to a billet crank because of the allegedly uninterrupted grain flow that can be obtained in the forging process. That might be true of some components, but with respect to crankshafts, the argument fails because of the large dislocations in the material that are necessary to move the crankpin and counterweight material from the center of the forging blank to the outer extremes of the part. The resulting grain structure in the typical V8 crank forging exhibits similar fractured grain properties to that of a machined billet. More than one crankshaft manufacturer has told me that there is no way that a forging from the commonly used steel alloy SAE-4340 (AMS-6414) would survive in one of today's Cup engines.

Some years ago, there was an effort at Cosworth to build a Formula One crankshaft by welding together various sections, which comprised the journals, webs and counterweights. The purported intent was to be better able to create exactly the shape and section of the various components, thereby reducing MMOI while achieving the same or better stiffness. While no one was willing to divulge details about the effort, it is rumored to have been run once or twice then abandoned due to the high cost and complexity compared to the measurable benefits.

In certain cases, there are benefits to the use of a built-up crankshaft. Because of the 'master-rod' mechanism necessary for the implementation of the radial piston engines that powered most aircraft until well into the second half of the 20th century, a bolted-together crankshaft configuration was used almost exclusively. **Figure 5** illustrates a typical two-row composite radial crankshaft and master-rod layout. The loose counterweights will be addressed later in this article.

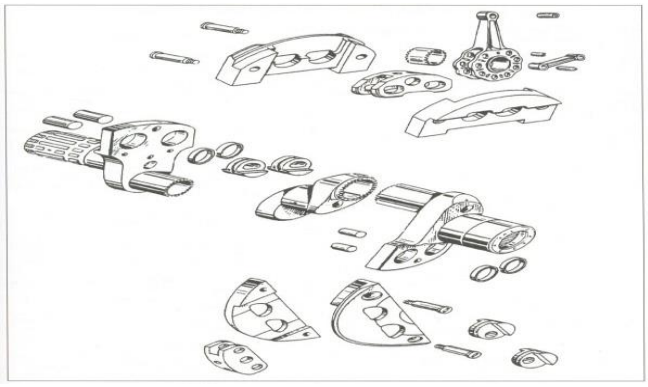


Figure:3

Built-up Radial Engine Crank

2) CRANKSHAFT MATERIALS:

The steel alloys typically used in high strength crankshafts have been selected for what each designer perceives as the most desirable combination of properties. **Figure 4** shows the nominal chemistries of the crankshaft alloys discussed here.

Medium-carbon steel alloys are composed of predominantly the element iron, and contain a small percentage of carbon (0.25% to 0.45%, described as '25 to 45 points' of carbon), along with combinations of several alloying elements, the mix of which has been carefully designed in order to produce specific qualities in the target alloy, including hardenability, nitridability, surface and core hardness, ultimate tensile strength, yield strength, endurance limit (fatigue strength), ductility, impact resistance, corrosion resistance, and temper-embrittlement resistance. The alloying elements typically used in these carbon steels are manganese, chromium, molybdenum, nickel, silicon, cobalt, vanadium, and sometimes aluminium and titanium. Each of those elements adds specific properties in a given material. The carbon content is the main determinant of the ultimate strength and hardness to which such an alloy can be heat treated.

Chemistry of Crankshaft Alloys
Nominal Percentages of Alloying Elements

Material	AMS	C	Mn	Cr	Ni	Mo	Si	V
4340	6414	0.40	0.75	0.82	1.85	0.25		
EN-30B		0.30	0.55	1.20	4.15	0.30	0.22	
4330-M	6427	0.30	0.85	0.90	1.80	0.45	0.30	0.07
32-CrMoV-13	6481	0.34	0.55	3.00	<0.30	0.90	0.25	0.28
300-M	6419	0.43	0.75	0.82	1.85	0.40	1.70	0.07
Key:	C = Carbon		Mn = Manganese			Cr = Chromium		
	Ni = Nickel		Mo = Molybdenum			Si = Silicon		
	V = Vanadium		AMS = Aircraft Material Spec Number					

Figure: 4

In addition to alloying elements, high strength steels are carefully refined so as to remove as many of the undesirable impurities as possible (sulfur, phosphorous, calcium, etc.) and to more tightly constrain the tolerances, which define the allowable variations in the percentage of alloying elements. The highest quality steels are usually specified and ordered by reference to their AMS number (Aircraft Material Specification). These specs tightly constrain the chemistry, and the required purity can often only be achieved by melting in a vacuum, then re-melting in a vacuum to further refine the metal. Typical vacuum-processing methods are VIM and VAR.

There are other ultra-high-strength steels that are not carbon steels. These steels, known as "maraging" steels, are refined so as to remove as much of the carbon as possible, and develop their extreme strength and fatigue properties as a by-product of the crystalline structures resulting from the large amounts of nickel (15% and up) and cobalt (6% and up) they contain. These steels can achieve extreme levels of strength and maintain excellent levels of impact resistance. As far as I could determine, maraging

alloys are not currently (2008) used for racing crankshafts but they have been used in certain extreme application conrods.

The material which is currently viewed as the ultra-extreme crankshaft alloy is a steel available from the French manufacturer Aubert & Duval, known as 32-CrMoV-13 or 32CDV13. It is a deep-nitriding alloy containing 300 points of chrome, developed in the mid-nineties specifically for aerospace bearing applications. It is available in three grades. GKH is the commercial purity and chemistry tolerance. GKH-W is the grade having higher purity (VAR) and tighter chemistry tolerance. GKH-YW is the extremely pure grade (VIM - VAR) and is said to cost twice as much per pound as the -W grade.

According to data supplied by Aubert & Duval, fatigue-tests of the -W and -YW grades, using samples of each grade heat treated to similar values of ultimate tensile strength, show consistently that the -YW grade achieves a dramatic improvement (over 22%) in fatigue strength compared to the -W grade, and the endurance limit is claimed to be just a bit short of the yield stress, which is truly amazing. I have been told that, because of the extreme stress levels on Formula One crankshafts, most of them use the -YW grade, while the lower stress levels of a Cup crank allow the successful use of the -W grade.

One well-known manufacturer (Chambon) has developed a process which allows the production of a deep case nitride layer in this alloy (almost 1.0 mm deep, as compared to the more typical 0.10 to 0.15 mm deep layer). They say this deeper case provides a far less sharp hardness gradient from the >60 HRc surface to the 40-45 HRc core, which improves the fatigue and impact properties of the steel. It says that its deep-case process requires several days in the nitriding ovens, but the depth allows finish-grinding after nitriding, using a very sophisticated process to remove the distortions which occurred during the nitriding soak.

No discussion of high-end crankshaft materials would be complete without mention of the ultra-high-strength alloy known as 300-M (AMS 6419). This alloy is a modification to the basic 4340 chemistry, in which a few more points of carbon are added (higher achievable hardness and strength), along with 170 points of silicon and 7 points of vanadium. The vanadium acts as a grain refiner, and the silicon enables the material to be tempered to very high strength (285 ksi) and fatigue properties, while retaining extremely good impact resistance and toughness.

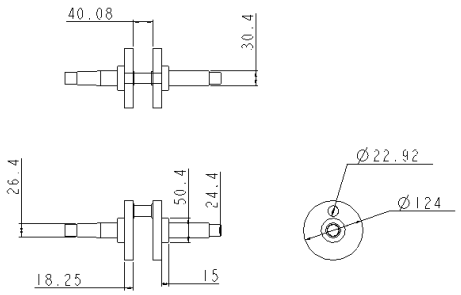
This material (300-M) is expensive and sometimes hard to get, since it is preferred for heavy aircraft landing gear components. It has been used by a few manufacturers for extreme duty crankshafts and conrods as well as high-shock aircraft components. However, several of the manufacturers I spoke with told me that they consider their favorite materials to be much better than 300-M for crankshaft applications.

3 DESIGN OF CRANKSHAFT:

SPECIFICATIONS:-

Lml freedom Bore diameter or cylinder bore = D=69.6mm Stroke = 82mm Explosion pressure gas pressure =P =10.936 N/mm2 Maximum torque = 200 N-m @1750-3000 rpm Design of crank shaft when the crank is at the dead centre We know that piston gas load Now the various parts of the crank shaft are design as discussed below: 1. Design of crank pin: let We have Takin 41585.951= Let us now the included bending stress in the crank pin We know that bending moment at the crank pin M = Section modules of the crank pin Z = Bending stress induced The induced bending stress is within in the permissible limits of 560 Mpa, therefore design of crank pin is safe. 2. Design of bearings Let Let us take thickness of the crank web t = 0.6 length of bearing WKT bending moment at centre of bearing M = =41585.951(0.75) =6133927.773 Bending moment (M) WKT 6133927.773= 3. Design of crank web Let w = width of the crank web in mm WKT bending moment on the crank web M = =41585.951(0.75)=2703086.815 Section modulus Z = Bending stress Total stress on the crank web = Total stress should not exceed permissible limit of 560 Mpa 560 = W = 17.390=20mm Design of shaft under the flywheel: Let First of all let us find the horizontal and vertical reactions at the bearings 1 and 2 Allowing for certain clearance the distance b = W+ 300+ 442mm and a = 0.75 WKT the horizontal reactions at bearings 1 and 2 due to the piston gas load () are Assuming The vertical reactions at bearings a1 and 2 ue to the weight of flywheel are W = weight of flywheel = 1.18kg=11.57N There is no belt tension the horizontal reactions due to the belt tension are neglected WKT horizontal bending moment at the flywheel location due to piston gas load = 3066964 There is no belt pull. There will be no horizontal bending moment due to the belt pull, i.e Total horizontal bending moment WKT vertical bending moment due to the flywheel weight Resultant bending moment = = 3066964.266 We know that bending moment 3066964.266.

2D DRAWING



INTRODUCTION TO CAD

Computer-aided design (**CAD**), also known as **computer-aided design and drafting** (**CADD**), is the use of computer technology for the process of design and design-documentation. Computer Aided Drafting describes the process of drafting with a computer. CADD software, or environments, provide the user with input-tools for the purpose of streamlining design processes; drafting, documentation, and manufacturing processes. CADD output is often in the form of electronic files for print or machining operations. The development of CADD-based software is in direct correlation with the processes it seeks to economize; industry-based software (construction, manufacturing, etc.) typically uses vector-based (linear) environments whereas graphic-based software utilizes raster-based (pixelated) environments.

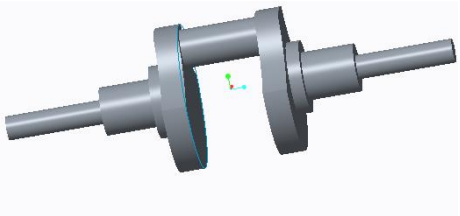
CADD environments often involve more than just shapes. As in the manual drafting of technical and engineering drawings, the output of CAD must convey information, such as materials, processes, dimensions, and tolerances, according to application-specific conventions.

CAD may be used to design curves and figures in two-dimensional (2D) space; or curves, surfaces, and solids in three-dimensional (3D) objects.

INTRODUCTION TO CREO

PTC CREO, formerly known as Pro/ENGINEER, is 3D modeling software used in mechanical engineering, design, manufacturing, and in CAD drafting service firms. It was one of the first 3D CAD modeling applications that used a rule-based parametric system. Using parameters, dimensions and features to capture the behavior of the product, it can optimize the development product as well as the design itself. The name was changed in 2010 from Pro/ENGINEER Wildfire to CREO. It was announced by the company who developed it, Parametric Technology Company (PTC), during the launch of its suite of design products that includes applications such as assembly modeling, 2D orthographic views for technical drawing, finite element analysis and more.

The time saved by using PTC CREO isn't the only advantage. It has many ways of saving costs. For instance, the cost of creating a new product can be lowered because the development process is shortened due to the automation of the generation of associative manufacturing and service deliverables. PTC also offers comprehensive training on how to use the software. This can save businesses by eliminating the need to hire new employees.



ADVANTAGES OF CREO PARAMETRIC SOFTWARE

1. Optimized for model-based enterprises
2. Increased engineer productivity
3. Better enabled concept design
4. Increased engineering capabilities
5. Increased manufacturing capabilities
6. Better simulation
7. Design capabilities for additive manufacturing

CREO parametric modules:

- Sketcher
- Part modeling
- Assembly
- Drafting

INTRODUCTION TO FEA

Finite element analysis is a method of solving, usually approximately, certain problems in engineering and science. It is used mainly for problems for which no exact solution, expressible in some mathematical form, is available. As such, it is a numerical rather than an analytical method. Methods of this type are needed because analytical methods cannot cope with the real, complicated problems that are met with in engineering. For example, engineering strength of materials or the mathematical theory of elasticity can be used to calculate analytically the stresses and strains in a bent beam, but neither will be very successful in

finding out what is happening in part of a car suspension system during cornering.

INTRODUCTION TO ANSYS

ANSYS Autodyn is computer simulation tool for simulating the response of materials to short duration severe loadings from impact, high pressure or explosions.

1. ANSYS Mechanical

ANSYS Mechanical is a finite element analysis tool for structural analysis, including linear, nonlinear and dynamic studies. This computer simulation product provides finite elements to model behavior, and supports material models and equation solvers for a wide range of mechanical design problems. ANSYS Mechanical also includes thermal analysis and coupled-physics capabilities involving acoustics, piezoelectric, thermal–structural and thermo-electric analysis.

2. Fluid Dynamics

ANSYS Fluent, CFD, CFX, FENSAP-ICE and related software are Computational Fluid Dynamics software tools used by engineers for design and analysis. These tools can simulate fluid flows in a virtual environment — for example, the fluid dynamics of ship hulls; gas turbine engines (including the compressors, combustion chamber, turbines and afterburners); aircraft aerodynamics; pumps, fans, HVAC systems, mixing vessels, hydro cyclones, vacuum cleaners, etc.

Definitions of Results obtained

Displacement - A vector quantity which refers to the distance which an object has moved in a given direction. It is measured as the length of a straight line between the initial and final positions of a body.

Von Mises Stress - The Von Mises criteria is a formula for combining these 3 stresses into an equivalent stress, which is then compared to the tensile stress of the material.

RESULTS

As per the analysis images

	Displacement (mm)	Von Mises Stress (N/mm^2)	Von-mises strain
Aluminum alloy 7475	0.004071	13.3622	0.191e-03
Aluminum alloy 6061	0.004154	13.6322	0.199e-03

Thermal analysis results table

	Nodal temperature(k)	Thermal gradient	Thermal flux
Aluminum alloy 7475	558	52.7603	7.20082
Aluminum alloy 6061	558	49.4276	8.5987

Manufacturing component

Casting Process

Consider the casting process at 645 B.C, the first traces of the Sand Molding was found. Now consider the state-of-the-art Electromagnetic casting process. Truly, the Casting process has traversed a long path and impacted human civilization for nearly five millennia. With technological advances, metal casting is playing a greater role in our everyday lives and is more essential than it has ever been.

Selecting the Right Metal Casting Process

- For any Metal Casting Process, selection of right alloy, size, shape, thickness, tolerance, texture, and weight, is very vital.
- Special requirements such as, magnetism, corrosion, stress distribution also influence the choice of the Metal Casting Process.
- Views of the Tooling Designer; Foundry / Machine House needs, customer's exact product requirements, and secondary operations like painting, must be taken care of before selecting the appropriate Metal Casting Process.
- Tool cost.
- Economics of machining versus process costs.
- Adequate protection / packaging, shipping constraints, regulations of the final components, weights and shelf life of protective coatings also play their part in the Metal Casting process.

Comparative Advantages, Disadvantages and Applications for Various Casting Methods:

Sand Casting		
Advantages	**Disadvantages**	**Recommended Application**
Least Expensive in small quantities (less than 100) Ferrous and non - ferrous metals may be cast Possible to cast very large parts. • Least expensive tooling	Dimensional accuracy inferior to other processes, requires larger tolerances Castings usually exceed calculated weight Surface finish of ferrous castings usually exceeds 125 RMS	Use when strength/weight ratio permits Tolerances, surface finish and low machining cost does not warrant a more expensive process
Permanent and Semi-permanent Mold Casting		
Less expensive than Investment or Die Castings Dimensional Tolerances closer than Sand Castings Castings are dense and pressure tight	Only non-ferrous metals may be cast by this process Less competitive with Sand Cast process when three or more sand cores are required Higher tooling cost than Sand Cast	Use when process recommended for parts subjected to hydrostatic pressure Ideal for parts having low profile, no cores and quantities in excess of 300
Plaster Cast		
Smooth "As Cast" finish (25 RMS) Closer dimensional tolerance than Sand Cast • Intricate shapes and fine details including thinner "As Cast" walls are possible • Large parts cost less to cast than by Investment process	More costly than Sand or Permanent Mold-Casting Limited number of sources Requires minimum of 1 deg. draft	Use when parts require smooth "As Cast" surface finish and closer tolerances than possible with Sand or Permanent Mold Processes
Investment Cast		
Close dimensional tolerance Complex shape, fine detail, intricate core sections and thin walls are possible Ferrous and non-ferrous metals may be cast As-Cast" finish (64 - 125 RMS)	Costs are higher than Sand, Permanent Mold or Plaster process Castings	Use when Complexity precludes use of Sand or Permanent Mold Castings The process cost is justified through savings in machining or brazing Weight savings justifies increased cost
Die Casting		
Good dimensional tolerances are possible	Economical only in very large quantities due to	Use when quantity of parts justifies the high tooling

CONCLUSION

In our project we have designed a crankshaft for a multi cylinder engine using theoretical calculations and modeled the crankshaft in parametric software creo. Pressure produced in the engine is also calculated.

Structural and modal analysis is done on the crankshaft to validate our design. Analysis is done for two materials aluminum alloy 6061 and Aluminum alloy7475.

By observing the stress values for both the materials, the analyzed stress values are less than their respective yield stress values. So our design is safe.

By comparing the stress results for both materials, it is less for Aluminum alloy7475 than aluminum alloy 6061.

So for our designed crankshaft, using Aluminum alloy 7475 is best.

REFERENCES

1. Design of machine member's vol.1&2 by R.S Kurmi, S.Chand publications.
2. C.M.Balamurugan.,R.Krishnaraj.,Dr.M.Sakthive.l, .Kanthave.l, DeepanMarudachalam M.G., R.Palani., 2011"Computer Aided Modeling and Optimization of Crankshaft, ''International Journal of Scientific & Engineering Research Volume 2, Issue 8 ISSN 2229-5518
3. Yu Ding and Xiaobo Li.,2011, " Crankshaft Strength Analysis of a Diesel Engine Using Finite Element Method," Asia-Pacific Power and Energy Engineering Conference
4. JianMeng., Yongqi Liu., Ruixiang Liu.,2011,"Finite Element Analysis of 4-Cylinder Diesel Crankshaft, " I.J. Image, Graphics and Signal Processing, 5, 22-29
5. Yu Gongzh.i, Yu Hongliang.,DuanShulin., 2011, "Crankshaft Dynamic Strength Analysis for Marine Diesel Engine," Third International Conference on Measuring Technology and MechatronicsAutomation

Heat Transfer Augmentation in Three Sides Dimple Roughened Solar Duct

Vikash Kumar[1], Meghavath Parsha[2]

[1,2] Assistant Professor, Department of Mechanical Engineering, Malla Reddy College of Engineering, Maisammaguda, Dhulapally, Kompally, Secunderabad-100, Telangana, India

Email: - (vikashism2012@gmail.com, parshamegavath@gmail.com)

Abstract

The present paper deals with the experimental results of heat transfer, friction factor and thermal efficiency of a novel type of three sides concave dimple roughened solar air heater under fully developed turbulent flow conditions. Three sides concave dimple roughened solar air heaters have higher values of heat transfer than those of one side concave dimple roughened solar air heaters in the range of 25-86% for varying relative roughness pitch and 21-81% for varying relative roughness height for the same values of operating parameters. The rise in friction factor of three sides roughened duct over one side roughened duct for varying relative roughness pitch and relative roughness height was found to be respectively as 11-34% and 15-41%. The maximum values of Nusselts number is obtained at relative roughness pitch of 12 and relative roughness height of 0.036. The maximum values of friction factor is obtained at relative roughness pitch of 8 and relative roughness height of 0.045.

Keywords: Solar air heater, Dimple shape, One side roughened duct, Relative roughness pitch, Relative roughness height, Three sides roughened duct.

1. Introduction

Solar air heaters follow a solar thermal technology in which the energy from the sun is captured by an absorbing medium and used to heat air. Solar air heating is a renewable energy heating technology used to heat or condition air for buildings or process heat applications. It is typically the most cost-effective out of all the solar technologies, especially in commercial and industrial applications, and it addresses the largest usage of building energy in heating climates as space heating and industrial process heating [1]. The value of heat transfer coefficient and heat capacity for air is low which reduces the heat transfer rate and thus increases the heat loss to the surroundings. Efforts have been made to improve the thermal efficiency by devising roughness in the form of wires, ribs, dimples baffles, fins, making the surface corrugated and packed bed and many more [2]. The low efficiency of SAH is due to less heat transfer between the collector and beneath flowing air, consequently raising collector's temperature, which leads in greater heat loss from the collector's surface to the nearby surroundings [3-7]. Literature of SAH reveals that the development of laminar sub-layer on heat exchanging surfaces and low heat capacity of air are responsible for less heat transfer resulting in lower thermal efficiency [8-9]. Even though the implication of roughness on heat exchanging surfaces results in an appreciable enhancement in heat transfer from the collector's surface to the under flowing air, the increase in frictional losses needs to be considered precisely. Heat transfer augmentation must be achieved at the cost of minimizing frictional losses. In order to dismantle the viscous sub-layer the core flow mustn't be disturbed. This is the reason the

selection of roughness geometry and orientation is given utmost importance [10-14].

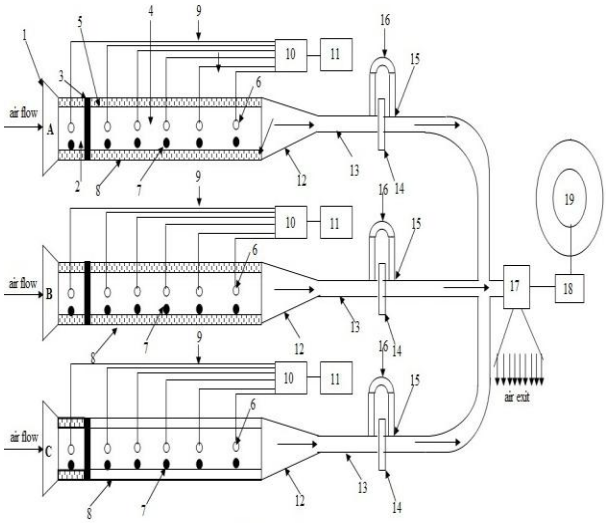
Prasad and Saini [9] provided roughness as transverse ribs of small height and explained the effects of p/e and e/D_h on heat transfer and friction factor. It was found that the maximum augmentation in Nu and f were 2.38 and 4.25 times respectively over the non roughened surface. Gupta et al. [15-16] studied the effect of, p/e, e/D_h, α and Re on the performance of an inclined wire roughened SAH and found that the maximum augmentation in Nu and f were at 60˚ angle of attack. Gupta et al also stated that in comparison to transverse wire, the inclined wire have more heat transfer augmentation due to development of the secondary flow accompanied with the destruction of the viscous sub-layer in the vicinity of the ribs attached to the roughened duct. Momin et al. [17] conducted experiments on V-shaped roughened ducts and discussed its effect on heat transfer and friction factor characteristics under a varying range of roughness geometries and flow parameters as Re between 2500-18000, e/D_h between 0.02-0.034 and constant p/e of 10. The maximum increase in Nusselts number and friction factor was found to be 2.30 and 2.83 times over non-roughened ducts. Varun et al. [18] provided roughness in the form of combination of transverse and inclined ribs and concluded that the geometry having p/e = 8 yielded maximum thermal efficiency.

There was a conclusive information in the literature that most of the roughness provided was limited to a single side (flow facing side) of the absorber plate. If roughness is provided to the side walls (2 nos) as well, they can participate in the heat transfer augmentation process accompanied by the slightest increase in pressure drop resulting in an appreciable enhancement in heat transfer [19-20]. The main objective of the present study is to provide roughness on the three sides (one top and two side walls) of the roughened duct and determine the augmentation in heat transfer and friction factor in compared to the single side roughened duct.

In the present investigation, roughness provided is in the form of concave dimple-shape on three sides and one side roughened duct under identical flow and geometrical conditions. The present paper projects the augmentation in Nusselts number and friction factor for three sides over one side roughened duct as a function of Reynolds number (Re), relative roughness pitch (p/e) and relative roughness height (e/D_h).

2. Experimental set-up and methodology

Experimentation under actual outdoor conditions has been performed on multiple sets of one and three sides roughened ducts containing dimple roughened absorber plates of varying roughness dimensions. The schematic diagram of the test setup is shown in Fig. 1. A 2130 mm x 630 mm wooden board of 25 mm thickness is used to formulate three ducts namely A, B and C. Each duct has a length of 2130 mm in which 630 mm serves as entry section and 1500 mm serves as test section. The entire duct design is based on ASHRAE Standard [21].



1. Trapezoidal shaped air inlet
2. Non-roughened duct section

3. Insulation between entry and test length
4. Insulation
5. Thermocouple
6. Thermometer
7. Glass covers
8. Copper wire
9. Selector switch
10. Digital voltmeter
11. Diverging section
12. Cylindrical pipe
13. Roughened duct section
14. Orifice-plate
15. Flange couplings
16. U-tube manometer
17. Blower
18. Motor
19. Variac

Fig.1. Schematic diagram of the experimental set-up

Once the flow is stabilized and the stagnation condition is achieved, readings for inlet and outlet air temperatures, plate temperatures, pressure drop across the duct and the orifice and solar insolation is recorded. Photograph of the experimental set-up is shown in Fig. 2

Fig.2. Photograph of the experimental set-up

The thermocouples were connected to a digital voltmeter indicating their output. The pressure drop across the orifice was measured with a U-tube manometer and along the duct was measured using multi-tube manometer having least count of 0.001 mm of H_2O. Pyranometer was used to measure the intensity of solar radiation and ambient temperature at different time intervals. The range of operating parameters is shown in Table. 1.

Table 1. Range of operating parameters

S. No.	Name of parameter	Symbolic representation	Range of operating parameter
1.	Reynolds number	Re	2500-12500
2.	Relative roughness pitch	p/e	8-15
3.	Relative roughness height	e/D_h	0.018-0.045
4.	Ambient temperature	T_∞	(24-44) ˚C
5.	Solar insolation	I	(720-960) W/m^2

3. Data Reduction

The recorded data from experimentation for plate and air temperatures under steady state conditions at a varying mass flow rate and heat flux was used to calculate the heat transfer from the absorber plate to the under flowing air. Using the pressure drop across the orifice plate, the prevailing mass flow rate across the duct is:

$$\dot{m} = C_d A_o \left[\frac{2\rho_a \Delta P_o \sin\theta}{1-\beta^4} \right]^{0.5} \quad (1)$$

Heat lost by the collector = Heat gained by the under flowing air

$$\text{i.e. } hA_p\left(T_{pm} - T_{fm}\right) = \dot{m}C_p\left(T_o - T_i\right) \quad (2)$$

$$\text{i.e. } h = \frac{\dot{m}C_p\left(T_o - T_i\right)}{A_p\left(T_{pm} - T_{fm}\right)} \quad (3)$$

Nusselts number is determined using heat convective heat transfer co-efficient as:

$$Nu = \frac{hD_h}{K} \tag{5}$$

The friction factor was determined for the test length 1500 mm using Darcy Weisbach equation as:

$$f = \frac{(\Delta P_d) D_h}{2\rho L V^2} \tag{6}$$

The uncertainty prevailing in the measurement of various parameters has been calculated following a simple procedure suggested by Klein and McClintock [22], the values of uncertainty involved in the measurement of Nusselts number and friction factor were found to be $\pm 3.28\%$ and $\pm 4.16\%$ respectively.

4. Validation

The experimental values of Nusselts number and friction factor for one side concave dimple roughened duct have been compared with those of a similar duct model adopted by Saini and Verma [23].

The Nusselts number and friction factor for one side dimple roughened duct is given by the eq. no. 9 and 10.

$$Nu_{1r} = 5.2\times10^{-4}(\mathrm{Re})^{1.27}\left(\frac{p}{e}\right)^{1.15}\left(\frac{e}{D_h}\right)^{0.0333} \times\left[\exp(-2.12)\left(\log\left(\frac{p}{e}\right)\right)^2\right] \times\left[\exp(-1.30)\left(\log\left(\frac{e}{D_h}\right)\right)^2\right] \tag{9}$$

$$f_{1r} = 0.642\,\mathrm{Re}^{-0.423}(p/e)^{-0.465}\times(e/D_h)^{-0.0214} \times\left[\exp(0.054)(\log(p/e))^2\right] \times\left[\exp(0.840)(\log(e/D_h))^2\right] \tag{10}$$

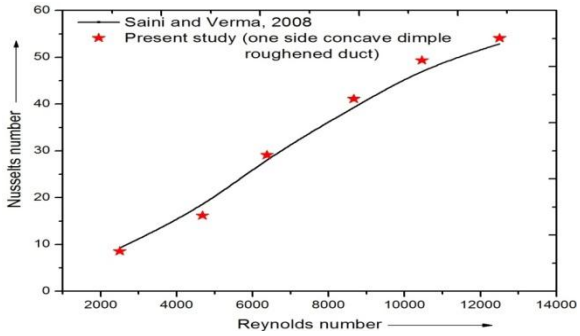


Fig. 7. Comparison of estimated and experimental values of 'Nu' for one side roughened duct

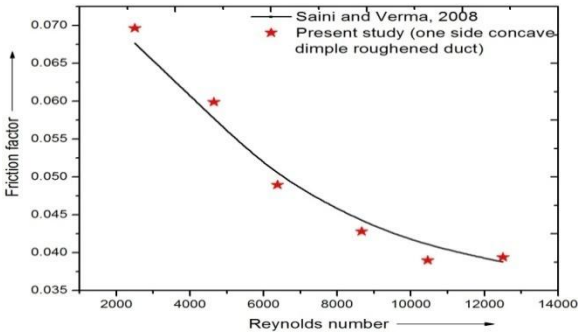


Fig. 8. Comparison of estimated and experimental values of 'f' for three sides roughened duct

The value for Nusselts number and friction factor for one side dimple roughened duct was found to compare well with the values so obtained from the correlations suggested by Saini and Verma. The mean deviation between the estimated and the experimental value was found to be following and the same has been shown in Fig.7 and 8:

- ±3.7% (Nusselts number)
- ±4.5% (Friction factor)

Based on the above comparison, reasonably good agreement between the experimental and estimated values of Nusselts number and friction factor guarantees the exactness of the information being gathered from this test setup.

5. Results and Discussions

The effects of dimple shape roughness element and the variation in the roughness parameter on heat transfer and fluid flow characteristics has been investigated and discussed. The values of the Nusselts number and friction factor for the three sides roughened ducts as a function of Reynolds number have been compared to those of one side roughened duct under identical experimental conditions.

5.1 Nusselts number

The augmentation in Nusselts number achieved as a result of providing artificial roughness in the form of concave dimple shape on the three sides roughened duct over one side roughened duct with an increasing Reynolds number for varying relative roughness pitch (p/e) and relative roughness height (e/D_h) is shown in Fig. 9 and Fig. 10.

Fig. 9 depicts that with an increase in p/e ratio, the Nusselts number increases, but only up to a relative roughness pitch of 12 beyond which it tends to decrease with an increase in relative roughness pitch. The maximum and minimum values of Nusselts number is obtained at the relative roughness pitch value of 12 and 8 respectively at a constant relative roughness height of 0.036 for the range of parameters investigated.

Fig. 10 shows the effect of relative roughness height on Nusselts number. The maximum heat transfer rate is achieved at the relative roughness height of 0.036.

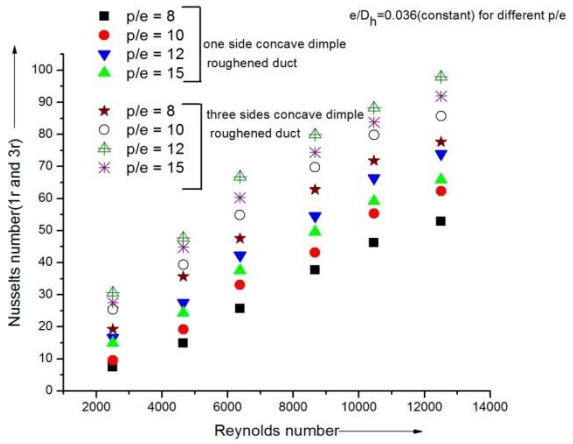


Fig.9. Effect of relative roughness pitch on 'Nu' for one and three sides roughened duct

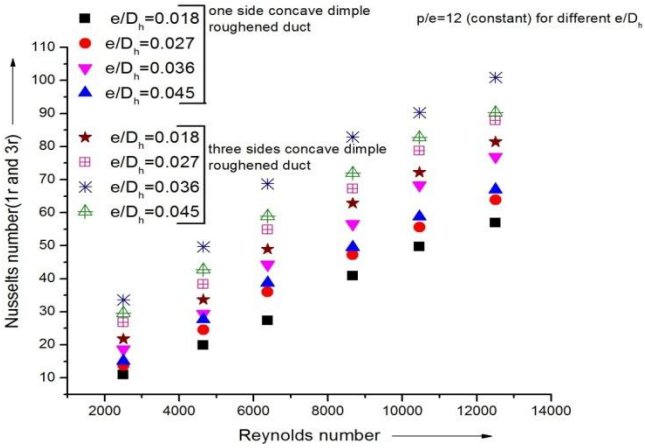


Fig.10. Effect of relative roughness height on 'Nu' for one and three sides roughened duct

5.2 Friction Factor

Artificially roughened SAHs are often characterized by the rise in pressure drop across the roughened duct, which results in an increment of friction co-efficient leading to a higher pumping power requirement. Numerous researchers have worked on different roughness geometry, trying to optimize the geometrical parameter to obtain a minimum rise in friction co-efficient. The effect of relative roughness pitch (p/e) and relative roughness height (e/D_h) on the friction factor with increasing Reynolds number is shown in Fig. 11 and 12 respectively.

Fig. 11 shows the variation in friction factor of three sides over one side roughened duct with Reynolds number with an increasing relative roughness pitch ratio. As the Reynolds number increases, the friction factor decreases monotonously. The three sides roughened duct requires more pumping power than one side roughened duct. For both the roughened ducts, friction factor decreases with increasing relative roughness pitch. The maximum and minimum friction factor for both three sides and one side roughened ducts is obtained at the relative roughness pitch values of 8 and 15 respectively.

Fig. 12 shows the variation in friction factor with Reynolds number with increasing relative roughness height ratio. It can be concluded that as the Reynolds number increases, the friction factor decreases with decreasing relative roughness pitch. Efforts should be made to optimize the geometrical parameter to achieve maximum heat transfer rate at a minimum rise in the co-efficient of friction. The maximum and minimum friction factor for both three sides and one side roughened duct is obtained at relative roughness height ratio of 0.045 and 0.018 respectively.

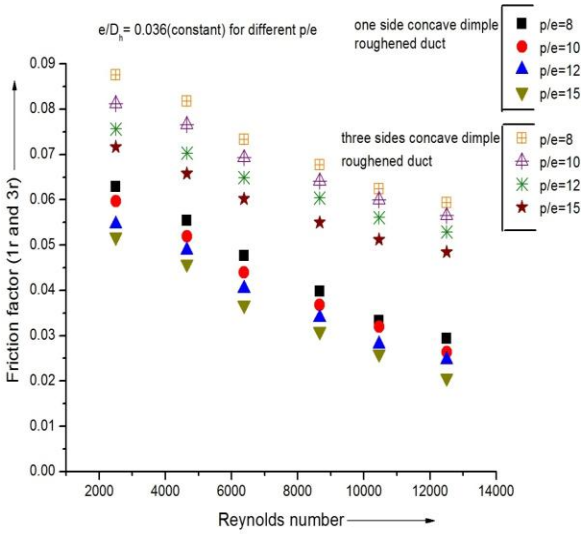


Fig.11. Effect of relative roughness pitch on 'f' for one and three sides roughened duct

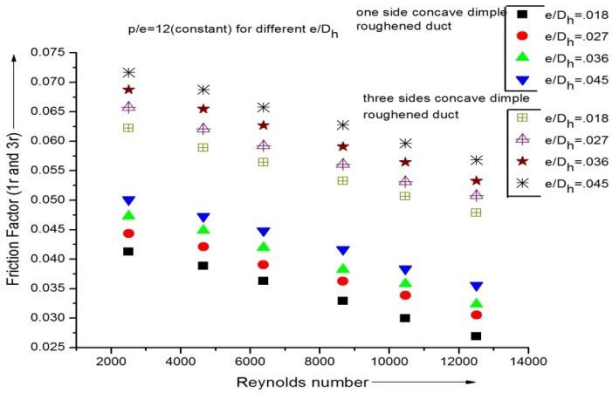


Fig.12. Effect of relative roughness height on 'f' for one and three sides roughened duct

6 Conclusions

Roughness geometry in the form of concave dimple resulted in an appreciable augmentation of heat transfer with an allowable rise in friction characteristics; following conclusions can be drawn:

- The maximum values of Nusselts number were obtained at relative roughness pitch of 12 and relative roughness height of 0.036.
- The augmentation in the Nusselts number of three sides roughened duct over one side roughened duct for varying relative roughness pitch in the range of parameters investigated was found to be 25-86%.
- The enhancement in the Nusselts number of three sides roughened duct over one side roughened duct for varying relative roughness height in the range of parameters investigated was found to be 21-81%.
- The maximum values of friction factor were obtained at relative

roughness pitch of 8 and relative roughness height of 0.045.

- The augmentation in friction factor of three sides roughened duct over one side roughened duct for varying relative roughness pitch and relative roughness height was found to be respectively as 11-34% and 15-41% in the range of parameters investigated.
- The average enhancement in the Nusselts number for three sides roughened duct over one side roughened duct for varying relative roughness pitch and relative roughness height was respectively found to be as 59% and 48% in the range of parameters investigated

References

1. Hsieh, J.S., Solar energy engineering. New Jersey: Prentice Hall; 1986.
2. Sharma, S.K., Kalamkar, V.R., Thermo-hydraulic performance analysis of solar air heaters having artificial roughness–A review Renewable and Sustainable Energy Reviews 41 (2015) 413–435.
3. Duffie, J.A., Beckman, W.A., Solar engineering thermal processes, 2nd ed. New York: John Wiley; 1991.
4. Sukhatme, S.P., Solar energy: Principles of thermal collection and storage, 9th Ed. New Delhi: Tata McGraw-Hill; 2003.
5. Dippreyy, D.F., Sabersky, R.H., Heat and momentum transfer in smooth and rough tubes at various Prandtl numbers. Int J Heat Mass Transf 1963; 6:329–53.
6. Webb, R.L., Eckort, E.R.G., Goldstein, K.J., Heat transfer and friction in tubes with repeated rib roughness. Int J Heat Mass Transf 1971; 14:601–17.
7. Firth, R.J., Meyer, L., A comparison of the heat transfer and friction factor performance of four different types of artificially roughened surface. Int J Heat Mass Transf 1983; 26(2):175–83.
8. Nguyen, T.M., Khodadadi, J.M., Vlachos, N.S., Laminar flow and conjugate heat transfer in rib roughened tubes. Numer Heat Transf Part A: Appl: Int J Comput Methodol 1989; 15:165–79.
9. Prasad, B.N., Saini, J.S., Effect of artificial roughness on heat transfer and friction factor in a solar air heater. Sol Energy 1988; 6:555–60.
10. Varun, Saini, R.P., Singal, S.K., A review of roughness geometry used in solar air heaters. Sol Energy 2007; 81:1340–50.
11. Ahn, S.W., The effect of roughness types on friction factor and heat transfer in roughened rectangular duct. Int Commun Heat Mass Transf 2001; 28:933–42.
12. Singh, A.P., Varun, Siddhartha, Effect of artificial roughness on heat transfer and friction characteristics having multiple arc shaped roughness element on the absorber plate. Sol Energy 2014; 105: 479–493.
13. Alam, T. and Kim, M.H., A critical review on artificial roughness provided in rectangular solar air. Renewable and Sustainable Energy Reviews 69 (2017) 387–400.
14. Hans, V.S., Saini, R.P. and. Saini, J.S. Performance of artificially roughened solar air heater-A review, Renewable and Sustainable Energy Reviews, 13, 2009, 1854-1869.
15. Gupta, D., Solanki, S.C., Saini, J.S., Heat and fluid flow in rectangular solar air heater ducts having transverse rib roughness on absorber plates. Sol Energy 1993; 51:31–37.
16. Gupta, D., Solanki, S.C., Saini, J.S., Thermohydraulic performance of solar air heaters with roughened absorber plates. Sol Energy 1997; 61:33–42.
17. Momin, A.M.E., Saini, J.S., Solanki, S.C., 2002. Heat transfer and friction in solar air heater duct with v-shaped rib roughness on absorber plate. International Journal of Heat Mass Transfer 45, 3383–3396.

18. Varun, Saini, R.P., Singal, S.K., Investigation of thermal performance of solar air heater having roughness elements as a combination of inclined and transverse ribs on the absorber plate. Renew Energy 2008; 33:1398–405.
19. Prasad, B.N., Behura, A.K. and Prasad, L., Fluid flow and heat transfer analysis for heat transfer enhancement in three sided artificially roughened solar air heater. Sol. Energy 105, 2014, 27–35.
20. Behura, A.K., Prasad, L. and Prasad, B.N., Heat transfer, friction factor and thermal performance of three sides artificially roughened solar air heaters. Solar Energy 130 (2016) 46–59.
21. ASHRAE, Standard, 93–97, 1977. Method of testing to determine the thermal performance of solar air heater. American society for heating, refrigeration and air conditioning engineering, New York, pp. 1–34.
22. Kline, S.J., McClintock FA. Describing uncertainties in single sample experiments. Mechanical Engineering 1953; 75: 3-8.
23. Saini, R.P., Verma, J., Heat transfer and friction factor correlations for a duct having dimple-shape artificial roughness for solar air heaters. Energy 2008; 33:1277–87.

Nomenclature

Symbol	Name of Parameter	Unit
L	Length of the roughened duct	[m]
W	Width of the roughened duct	[m]
H	Height of the roughened duct	[m]
$\dot{m}$	Mass flow rate	[kg/s]
C_p	Specific heat capacity	[J/kgK]
T_o	Air outlet temperature	[˚C]
T_i	Air inlet temperature	[˚C]
T_∞	Ambient air temperature	[˚C]
Q_u	Useful heat gain	[J]
C_d	Coefficient of discharge	
A_o	Area of orifice plate	[m^2]
A_p	Area of absorber plate	[m^2]
h	Convective heat transfer coefficient	[W/m^2K]
Nu	Nusselts number	
f	Friction factor	
T_{pm}	Plate mean temperature	[˚C]
T_{fm}	Fluid mean temperature	[˚C]
d	Pipe diameter	[m]
D	Orifice plate diameter	[m]
D_h	Hydraulic diameter	[m]
I	Solar insolation	[W/m^2]
p/e	Relative roughness pitch	
e/D_h	Relative roughness height	

Greek symbols

β	Ratio of pipe diameter to orifice diameter	
η	Thermal efficiency	[%]
θ	Inclination of U-tube manometer	[˚]
ρ	density	[Kg/m^3]

Demand of Plastic Fuel in India

First S. Afroz,, *M.Tech (Thermal)*, Second S. Ravi Sagar, *M.Tech (Thermal)*

***Abstract*—India Facing a bigger challenge today of recession in Public sector, Indian economy slowing down. The Major sector which is highly affected is FUEL. We all complain about the rising fuel prices, while the government is trying to promote biofuel across the country, many commuters have also started opting for electric mobility, which is costing them half the price of Petrol and Diesel Vehicle. At this point of time we should rethink about Fuel availability from non-biodegrable waste. If the plastic waste converted into energy our country will be able to solve the demand of power and plastic waste management. Plastic remain under the Ground for 500 years, which leads to the contamination of soil and thus pollute the Environment. Over the Years several waste to wealth mechanism have been adopted to recycle and reuse plastic in innovative ways. One such trends has been the conversion of Plastic waste to fuel and making it usable for both domestic and industrial purpose. Daily Generation of over 15,000 tonnes of plastic, the prospect to conversion to fuel are abundant, provided there is enough Infrastructure available. The current Union government also addresses the issues of plastic waste and ways with which the problem could be dealt with.**

***Index Terms*—Waste Management, Fuel, power, energy, wealth mechanism**

I. INTRODUCTION

we all are aware about the R's Recycle, Reuse and Regenerate. In the case of plastic which is non-biodegradable and very harmful for environment and major solid waste for municipal and Industries, due to lack of integrated solid waste management, most of the plastic waste neither collected properly nor disposed of in appropriate manner to avoid its negative impact on environment. Worldwide Plastic production is predicted to increase by 3.8% every year until 2030.Flexible non-recycle plastic material are used in an increasing range of application like packaging,3D printing and construction. We need to expand our range of option for keeping this plastic waste out of landfill. One potential approach is "plastic to energy", which unlock

chemical energy stored in waste plastic and uses it to create fuel. Countries like Japan, Germany and united states have already implemented the plastic to fuel conversion process with much success. These three is also been successful in creating business model out of the conversion process, resulting in a conversion model becoming a profitable business one. Though India still has long way to go in terms of adopting plastic to fuel as a business model, discoveries are being made to convert plastic to usable fuel.

II. WORK AHEAD FOR FUEL GENERATION FROM PLASTIC

Few Indian and company have taken initiative in this regard for plastic waste management and power generation to fulfil our demand of economic growth and to protect the environment from health hazard.

A. IIP(Indian Institute of Petrolium)

A constituent Laboratory of the council of scientific and Industrial Research (CSIR) in 2014, developed a unique process of converting plastic waste like polyethylene and polypropylene, both together accounting for 60 percent of plastic waste, can be converted to either gasoline or diesel. The technology can convert 1 kg of plastic to 750 ml of automotive grade gasoline. Due to nearly nil presence of Sulphur in the produced fuel, IIP'S plastic converted fuel is pure and meets the EURO-III standard. IIP also stated that a vehicle using this fuel would be able to run at least two Km more per liter. The technology was developed by IIP after nearly a decade of research in hope of commercializing it for Industrial usage.

B. Mechanical Engineer and Professor (Satish Kumar)

Mr. Satish Kumar has come up with a three-step process called plastic pyrolysis to convert plastic into fuel. Today, the fuel is being sold to local industries for as less as Rs. 40 per liter. He has registered the company with ministry of micro, small and medium Enterprises (MSME). Since 2016, he has converted 50 tonnes of unrecyclable plastic into fuel. At present, his company is recycling 200 200kg o plastic and producing 200 liters of petrol out of it every day. "The process helps recycle plastic into diesel, aviation fuel, and petrol. About 500 kg of non-recyclable plastic can produce 400 liters of fuel. It doesn't pollute the air as the process happens in a vacuum." However, using this fuel for vehicles is yet to be tested.

III. METHODOLOGY INVOLVED IN CONVERSION OF PLASTIC INTO FUEL

There are several processes of plastic conversion into fuels. Out of which two are Gasification and pyrolysis.

Gasification involves heating the waste plastic with air or steam, to produce a valuable industrial gas mixture called "synthesis gas", or syngas. This can then be used to produce diesel and petrol or burned directly in boilers to generate electricity.

Pyrolysis is generally defined as the controlled heating of a material in the absence of oxygen. In plastic Pyrolysis, the

macromolecular structure of polymer is broken down into smaller molecules and sometimes monomer units. Further degradation of these subsequent molecule depends on a no. of different condition including temperature, residence time, presence of catalyst. Accordingly, the reaction will be thermal and catalytic pyrolysis. Plastic waste is continuously treated in a cylindrical chamber. The plastic is pyrolyzed at 300 C -500 C.

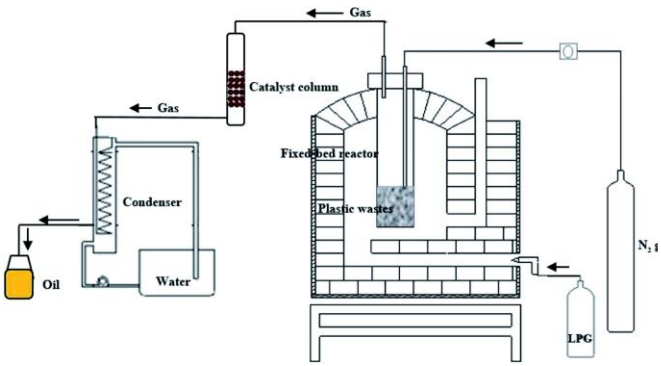


Fig. 1 Schematic diagram of bench scale pyrolysis unit.

A. Pyrolysis Process.

A bench scale fixed-bed pyrolysis stainless steel batch reactor (diameter 35 cm, length 60 cm, maximum capacity 5 kg) was used for production of oil from plastic packaging wastes (Fig. 1). 1 kg of plastic wastes was loaded into the reactor for each pyrolysis reaction. The reactor was heated externally by liquid propane gas to the required operating temperature at heating rates of 10, 15, 20 and 25 °C min−1. Reaction was carried out at 500 °C for 10 minute under a flow of nitrogen purge gas. The pyrolysis gas was passed through catalyst column at a ratio of 0.05, 0.1, 0.15 and 0.2 by weight of catalyst to plastic. The catalyst column made from glass (diameter 5 cm, length 100) and load with catalyst in pellet form. Pellets are formed on compression of the catalyst powder with hydraulic press at pressure of 50 bar, the resulting pellet have a diameter of 1 cm and length of 1.5 cm. No binder was used in the formation of the pellets. Fig. S1 in the ESI† demonstrates the bentonite clay, hydraulic press machine, palletization block and catalyst in pellet form. The resulting gas products were collected *via* water cooled condenser. The oil yield was determined based on the initial mass of plastic waste.

TABLE I

Resin	Conversion Rate
PET	30%
HDPE	Data not Available
PVC	30%
LDPE	70%
PP	50-60%
PS	80-85%

IV. Result and Discussion

The solution of plastics-to-fuel holds promise in not only curbing such pervasive pollution but also providing a significant economic benefit to regions. The American Chemistry Council estimates plastic-to-fuel facilities in the US alone would create nearly 39,000 jobs and almost $9bn in economic output, making the global market potential of such an industry huge.

Plastic-derived fuels are also capable of producing a cleaner burning fuel than traditional sources due to their low Sulphur content, considering the majority of developing nations use Sulphur-heavy diesel. Though slow, progress is being made on the waste to fuel conversion front in India. Unfortunately, despite the big stride in waste management, extensive setting up of waste to fuel plants across the country is still awaited. The technologies employed to convert plastic waste to fuel are not complicated to replicate, and if done so on a large scale, will only help in addressing the growing issue of India's plastic waste. Like other country such as Australia we could create direct job opportunities in plastic conversion plant, and indirect jobs around installation, maintenance and distribution of energy and fuels. We might even see jobs in R&D to explore other waste conversion technologies. This could be a better and upcoming job opportunities for Indians.

PTF technologies can address a critical fraction of the plastic waste stream that has been historically difficult to reutilize, such as LDPE, PP and PS, preventing littering and the landfill disposal of end-of-life plastics. When PTF technologies target end-of-life plastics that are not easily or economically absorbed in recycling markets, they are considered complementary to recycling and existing waste hierarchies. The development of PTF infrastructure can also:

- Create green indirect and direct jobs,
- Divert end-of-life plastics from landfill disposal, extending the lifespan of existing disposal sites and prolonging the siting and construction of new ones,
- Create local demand for low-value plastics that can find their way into streets, streams and the ocean,
- Produce a local source of synthetic crude oil and/or refined fuels to displace fossil fuel derived imports, and
- Reduce air pollution in many parts of the world by substituting low or ultra-low sulfur content fuels for high-sulfur content fuels.

References

[1] Matesh Basappa Khot. (2017, 09). Plastic Waste into Fuel Using Pyrolysis Process. *Journal.* [IRJET]. *volume (04),* 342. Available: www.irjet.net

[2] By Think Change Indiar. (2019,06). Think Change India.

*[3] :*Muxina Konarova, university of Queensland. (2014,05). If we can't recycle it , why not turn our waste plastic into fuel?

[4] Ocean Recovery Alliances.(2015,06). 2015 Plastic To Fuel Project Developer Guide. Available: www.oceanreco.org

[5] Saptrishi Dutta (2017, 08). Recycling Plastic In India: Converting Plastic Waste To Fuel, The Unrealized Potential.

Development of Automatic Gear Changing Mechanism

P.Nithin[1] E.Prashanth[2]
M.TECH
[1,2]Department of Mechanical Engineering MRCE

Abstract— In the current world of vehicle, gear changing system are manually and automatic. Gear changing system is main in automobile to vary the speed ,Manual gear changing system is less cost than automatic gear shifts system but manually gear shifting system is difficult to understand for people or it take physical effort to change gear. To beat this disadvantage we try to apply touch screen based automatic gear changing system. In automatic gear changing system by touch on touch screen panel gear is shift. By applying this gear changing system it gives cost reduction in compare of automatic gear shift system and flexible or simple then manually gear shift system, and by applying this automatic gear changing system on automobile everyone can drive car easily. The purpose of this research is to reduce physical effort of human being and they can think only in driving and avoid the accidents. In present report we studied literature review baised on this review we define transmission system and gear shifting mechanism to modify a manual gear changing mechanism.

Key words: Actuator, Micro-controller, - 4.3- Touch screen LCD, Four wheeler Engine, Manual gear box, Clutch plate

I. INTRODUCTION

A. Detailed Description Of Problem

Today use of automobile is growing rapidly and to survive in the market is quite difficult with previous technologies. So to survive in the market new technologies are mandatory. The technology must be flexible enough to undertake changes. The technology must continue to grow. During our study on the recent gear shifting mechanism we find out many problems in present system.

In the recent trend there are two existing gear shifting system consists 1) manual gear shifting 2) automatic gear shifting. Here are the problems of existing system are explained below:

1) **Manual Gear Shifting Mechanism:**

The problems in manual gear shifting mechanism are described below

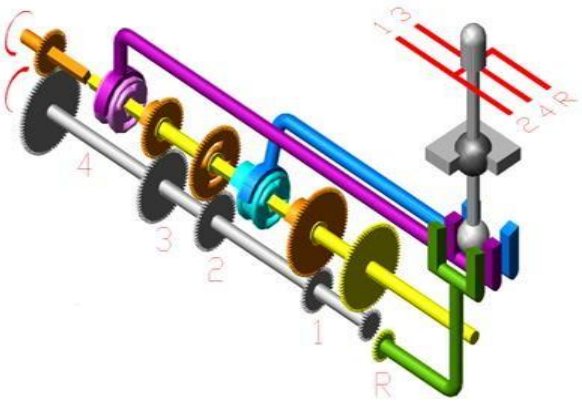


Fig. 1: Manual gear shift technology.

– Required More Human Effort

In manual gear shifting mechanism, the gear can be change with the help lever. To provide motion of lever human effort is required. Without applied human effort gear cannot be change, so sometime it's burdensome for car driver.

– Transmission Error

In manual shifting mechanism. The gear is changed by the driver so that it may be chance to shifting wrong gear instead of true gear. So in manual transmission system the chance of transmission error is high.

– Lack Of Response

We notice that while driving the car hesitates to refuse into gear. Manual transmissions can have the lack of response, so that car cannot run fast as per required in that gear.

– Humming Sound

Manual transmission will emit the sound while changing the gear. This sound is louder due to mechanical operation. The clucking sound is always done when we shift gear within transmission.

– Complex System

The manual transmission system is very complex because it is consist of complex mechanism like gear, lever operation, clutch engage & disengage mechanism.

– Gears Slipping

If the transmission system is immediately slipping in out of gear while driving, then chance of accident is high. So it is very risky. The reason of this problem due to hold the gears.

– Difficult To Understand

In manual transmission system the gear can be changed manually. For change gear we must be change lever operation so that it is difficult to remember &understand.

2) **Automatic Gear changing Mechanism**

There are following problems which occur in automatic gear shifting mechanism.

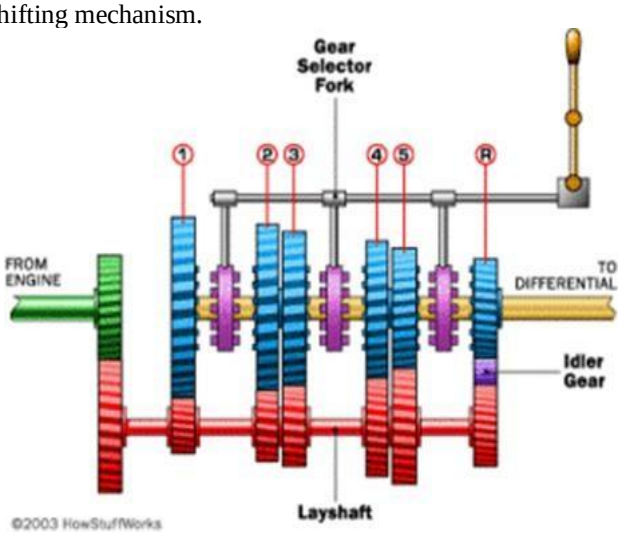


Fig. 2: Automatic gear shift technology.

– High Cost

The automatic transmission system is very costly due to its functioning.

– Less Efficiency

Automatic gear shifting mechanism have a low efficiency compare to manual gear shifting mechanism.

– Gear Slipping

As per mentioned in manual transmission problem gear slipping problem is also occurred in automatic transmission system.

– Grinding Or Shaking

Automatic transmission act a little differently, in automatic transmission system will take some time to wiggle into gear at first instead of typical smooth gear shifting.

– Delayed Engagement

Delayed engagement is one type of slip in this transmission. The clutch & brake do not operate instantly; due to this reason delayed engagement is occurred in automatic transmission system.

The purpose of this research is to reduce people physical effort and they present only in driving and prevent from accidents. In the present world of automobile, gear shifting system are manually and automatic.

Gear shifting system is important in automobile to vary the speed. So automatic gear shifting system is costly than manually gear shifts system and manually gear shifting system is difficult to understand for some people or it take physical effort to change gear.

To overcome this disadvantage we apply touch screen based automatic gear shifting system. In that system by touch on touch screen panel gear is shift. By applying this gear shifting system it gives cost reduction in compare of automatic gear shift system and flexible or simple then manually gear shift system. And by applying this system on automobile it's easier to drive car for everyone. It also reduces the possibility of transmission error of manual gear system. The following are the objective of touch screen based gear shifting mechanism.

1) Low Cost:

This system is less costly compared to automatic gear transmission system but it is costly compare to manual gear transmission system.

2) Flexibility:

This system is flexible compare to the manual transmission system so that it's beneficial to use this system compare to the manual gear shifting mechanism.

3) Easy to operate:

This is semi-automatic shifting mechanism in which touch screen are used so that gear can be shift with the help of pressing number so it is easy to operate.

4) Less human effort:

With the help of touch screen based gear shifting mechanism, no effort is required to change gear. Gear can change just with touch the appropriate number on the screen, so no effort is required

5) High efficiency:

The efficiency of semi-automatic gear shifting mechanism up to 85-90% ,whereas automatic gear shifting mechanism have a efficiency only 75%, so it is very reliable compare to other system.

6) Less transmission error:

Transmission error is occurring due to the improper gear at the certain speed. This is eliminating in this project because we provide indicator for wrong gear at specific speed.

7) Additional Outcomes:

– This project will save time and expenses of user.
– Easy to understand.
– Prevent from accident
– No emit sound while changing gear
– Applicable for all class

II. LITERATURE REVIEW

The literature studied provides so far insight information or new way to carry out this research. An attempt has been made to present the finding research paper on Gear Shift Mechanism. MT Gear Shifting Behavior indicated on manual transmission gear shift mechanism force required to shift gear he got result that max force required to shift gear back to one is 17.88N at shifting rod angle 12.86 °."A Brief Review of Transmission in Automobiles" revealed that the engine provides its highest torque outputs approximately in the middle of its range, while often the greatest torque is required when the vehicle is moving from rest or traveling slowly. "A Brief Review of Transmission in Automobiles" revealed that the engine provides its highest torque outputs approximately in the middle of its range, while often the greatest torque is required when the vehicle is moving from rest or traveling slowly. "Development of Actuator Control Strategy for DC Motor Controlled Automated Manual Transmission (AMT)" developed that Actuator control strategy for automated manual transmission (AMT) which uses electro mechanical Dc motor controlled linear actuators. Develop a strategy for deciding the operation of actuator. "Design Proposals for Low Cost Automated Manual Transmission (AMT)"analyzed that different automated manual transmission (AMT) system for automobile vehicle and developed a low cost design proposal for AMT system. In house designed actuators will require much time to prove the functionality according to our requirement and would demand increase in the development time of the project.) "A Literature Review on Automated Manual Transmission (AMT)" analyzed that transmission based actuator (TBA) uses multi-speed transmissions such that heavy, high-torque motors can be traded for high-speed, reduced mass motor-transmission combinations.) "A Literature Review on Automated Manual Transmission (AMT) analyzed that transmission based actuator (TBA) uses multi-speed transmissions such that heavy, high-torque motors can be traded for high-speed, reduced mass motor-transmission combinations.

III. CONCLUSION

The various research works shows that the system can be with hydraulic actuators or electrical based actuators. Actuators are basically used for gear shift actuations. It has the advantage of lower weight and higher efficiency with respect to other transmission system. Moreover, since AMT is directly derived from manual transmission with the integration of actuators into existing devices, the cost of this system is very less compare to Automatic transmission system.

IV. FUTURE SCOPE

With the help of touch screen people are easily applied gear without more effort. Touch screen can also be used in two wheel vehicle. Shifting of the gear system is more flexible and more reliable. An overall strategy aiming to the improvement of the gear shift quality should take into account the reduction of shifting time, the minimization of mean vehicle deceleration due to traction loss and the minimization of vehicle and driveline oscillations due to variation of transmitted torque. Existing transmission system can be converted into this transmission system. Here, the gear stick is to be replaced by touch screen for the driver to interact with the actuator, so that gear can be applied according to requirement. Future work should basically focus on low cost system with optimized control strategy.

REFERENCES

[1] Tian Shaopeng (2011) a study on "MT Gear Shifting Behavior" CISME Communications in Information Science and Management Engineering ,Vol.1 No.2 2011 PP.32-36 www.jc-isme.org

[2] Ranbir Singh (2012) "A Brief Review of Transmission in Automobiles" IJESRT- international journal of engineering sciences & research technology, Vol.1, issue 2: Page No.D222-D226, July-August (2012).

[3] Chinmay Kirtane (2013) "Gear Shift Schedule Optimization and Drive Line Modeling for Automatic Transmission" 1st International and 16th National Conference on Machines and Mechanisms (iNaCoMM2013), IIT Roorkee, India.

[4] Suraj S Raut (2014) "Automatic Transmission Gearbox with Centrifugal Clutches" IJESRT- international journal of engineering sciences & research technology, Volume: 03 Issue: 10 Oct-2014.

[5] Hiroshi Kuroiwa (2004) study on "Next-generation Fuel-efficient Automated Manual Transmission" Hitachi Review Vol.53 (2004).

EFFECT OF WELDING SPEED AND GROOVE ANGLE ON STRENGTH OF BUTT WELD JOINT USING TIG WELDING

MOHD ASLAM ALI
M Tech (Thermal Engineering)
Malla Reedy college of Engineering,Mysammaguda
(Affiliated to jntuh)
Hyderabad,Telangana

ABSTRACT

Welding is the metal joining process in which two or more metal having same material or different can be joined by heating to a plastic state .It is mostly used for joining metals in process industry, in fabrication, maintenance, repair of parts and structures. The metal plates and pipes used in process industry and they have welding strength as their important parameter.

In this thesis, the welding speed and geometry to find out tensile and impact strength in case of butt weld joint will be done. For V-groove geometry different models of plate with various included angles from 35^0, 45^0, 50^0will be made from structural steel (A633 Grade E).Currently different welding speeds are used in precision welding applications such as nuclear reactor pressure vessels, boilers etc. where welding accuracy as well as quality with strength is an important parameter. So in this project experimentation will be done on different welding speed such as 0.4 cm/sec, 0.8 cm/sec and 1.20cm/sec to prepare a V-groove butt weld joint. Generally the V-groove geometry with included angle up to 60^0 is in use.

Keywords: Welding Speed, Weld Joint, Tig Welding

CHAPTER-1

INTRODUCTION

Welding is, at its center, merely the way of bonding 2 objects of metallic. Whereas there are opportunity approaches in which to affix metallic (riveting, brazing and bonding, as an example), attachment has turn out to be the strategy of selection for its electricity, potency and flexibility.

There are loads of completely special attachment methods, and a number of ar being unreal all of the time. Some methods use warmth to generally melt 2 objects of metal along, commonly including a "filler steel" into the joint to behave as a binding agent. Opportunity methods deem pressure to bind metal along, and still others use a mixture of each heat and stress. No longer like bonding and brazing, anywhere the metal gadgets being joined stay unaltered, the approach of attachment continually adjustments the work items.

This may appear to be a trivial reason, however it is certainly vital to know-how why attachment produces such robust bonds. In the approaches of soldering and brazing, portions of steel arejoined with the aid of introducing a third material (with a decrease melting factor) into the mixture. Melting this 0.33 material among the surfaces of the unique portions binds the portions together. The bond, but, is handiest as robust as the becoming a member of material. Welding, then again, cuts out the intermediary and joins the original portions immediately to each other. The result is a strong, cohesive bond that's regularly as strong as the fabric itself.

As for substances, a few ar plenty of easier to weld than others. Steel may be an fantastic choice because of its power, affordability and weldability. As a rule, the more potent the metallic, the harder it's to weld. Consequently, many metallic alloys had been developed with attachment in thoughts. Of path, actually any metallic will be welded, in addition to forged iron, bronze, steel element and even metallic, although the latter wishes a extremely covered atmosphere as a result of the

metal is consequently reactive.

Whatever you are welding, bear in mind: protection first. If you have ever seen welding in man or woman, you could testify to the blinding brightness the system creates. Looking directly at a weld web page without protection can produce what is referred to as arc eye, a painful irritation of the cornea that feels like getting sand on your eye. No surprise that an awesome welder's mask is a prerequisite for any welding outfit.

Welding mask are available many patterns. The handiest ones have a darkened panel that the welder seems thru even as welding. More superior masks automobile-darken as the welding website receives brighter. In addition to mind-blowing brightness, welding can generate temperatures of up 10,000 ranges F (five,538 ranges Celsius) and showers of sparks, making heavy-obligation gloves and a long-sleeve shirt essential.

Lastly, proper ventilation is vital, relying on the welding approach. Welders may be exposed to dangerous substances along with lead, mercury and carbon monoxide. Vent hoods can prevent fumes from collecting within the workspace.

THE PROCESS OF WELDING

Most welding carried out nowadays falls into one in all two classes: arc welding- and torch welding. Arc welding use-s an electrical arc to soften the paintings materials in addition to filler cloth (occasionally referred to as the welding rod) for welding joints. Arc welding includes attaching a grounding cord to the welding material or other steel surface. Another twine referred to as an electrode lead is located at the material to be welded. Once that lead is pulled away from the material, an electric powered arc is generated. It's a little like the sparks you see while pulling jumper cables off a car battery. The arc then melts the work portions together with the filler fabric that allows to sign up for the pieces.

CLASSIFIATION OF WELDING PROCESSES

Welding technique can be labeled into distinct classes depending upon the following

criteria :

(a) It may be classified as fussion welding or strain welding depending upon at the application of heat. If software of warmth isn't always required, it's far referred to as stress welding.

(b) In case of fusion welding it may categorized low temperature welding and high temperature welding. When heat is generated to expand low temperature it's far called low temperature welding like soldering and brazing. Other fusion welding methods are high temperature welding strategies.

(c) Fusion welding can also be categorised on the idea of method of warmth era like gasoline welding, electric powered arc welding, resistance welding, thermit welding, and many others.

(d) On the basis of the kind of joint produced it is able to be categorised as butt welding, seam welding, spot welding, lap joint welding, and so on.

Each of the above sort of welding can be similarly labeled relying on other micro stage traits.

Gas Welding

It is a fusion welding wherein sturdy gas flame is used to generate warmth and lift temperature of metal pieces localized on the place wherein joint is to be made. In this welding metallic pieces to be joined are heated. The metallic for this reason melted starts off evolved flowing along the rims in which joint is to be made. A filler metallic may also be added to the flowing molten steel to top off thehollow space at the edges. The hollow space filed with molten metal is allowed to solidify to get the strong joint. Different combos of gases may be used to obtain a heating flame.

The famous fuel mixtures are oxy-hydrogen combination, oxygen-acetylene, and so forth. Distinct blending proportion of two gases in a mixture can generate exceptional sorts of flames with distinct characteristics.

Oxy-Acetylene Welding

Oxy-acetylene welding can used for welding of wide variety of metals and alloys. Acetylene mixed with oxygen while burnt below a managed surroundings produces large quantity of warmth giving higher temperature upward thrust. This burning additionally produces carbon dioxide which facilitates in stopping oxidation of metals being welded. Highest temperature that may be produced by this welding is 3200oC. The chemical reaction worried in burning of acetylene is

$2C_2H_2 + 5O_2 = 4CO_2 + 2H_2O + Heat$

on the basis of supply strain of gases oxy-acetylene welding is labeled as high stress welding on this device both gases oxygen and acetylene furnished to welding quarter are excessive strain from their respective excessive stress cylinders. The different one is low stress welding in which oxygen is furnished from excessive pressure cylinder but acetylene is generated by means of the action of water on calcium carbide and provided at low stress. In this example high strain supply of oxygen pulls acetylene at the welding zone.

An evaluation can be drawn among low strain and excessive stress welding. High strain welding equipment is handy, substances pure acetylene at constant strain, with better control and coffee fees in comparison to low stress welding.

CHAPTER-2

LITERATURE REVEIW

1.Effect of welding geometry parameter on hardness for aisi304 tig.

Welding is an area wherein technological tendencies out match the tendencies in its technology base that's generally driven with the aid of the outstanding industrial demand for welded structure. Reliability, Reproducibility and Viability necessities are forcing Technologists to take a look at weld defects consisting of distortion, warm cracking, in a systematic and logical technique than on experimental basis.

2. A overview paper on impact of welding pace and groove angle on Strength of butt weld joint the usage of tig welding.

Welding is most critical operation in any enterprise. It is crucial to optimize the diverse parameters of welding process in order that we can achieve the reliability, productivity and great of the goods. So industries are forcing the engineers to take a look at the welding manner parameters including electrodes, inert fuel, present day, voltage and so forth. The objective of any industry is manufacturing of excessive quality merchandise at low fee and increase the manufacturing fee. TIG welding system is versatile and normally used operation for joining of materials with the software of warmth and /or strain or fillet material to increase the production with much less time and price.

METHODOLOGY

Objective of the work

In this thesis, materials V-groove geometry distinct models of plate with diverse included angles from 350, 450, 500will be crafted from structural metal (A633 Grade E).Currently special welding speeds are including zero.Four cm/sec, zero.Eight cm/sec and 1.20cm/sec to put together a V-groove butt weld joint.

CHAPTER-3

EXPERIMENTAL PROCEDURE

In this thesis, experiments are made to understand the effect of TIG welding parameters welding speed and groove angle on output parameters such as hardness of welding, tensile strength of welding.For the experiment, welding parameters selected are shown in table.

The welding current and electrodes considered are

PROCESS PARAMETERS	LE VEL1	LEVEL2	LEVEL3
WELDING SPEED (cm/s)	0.4	0.8	1.20
GROOVE ANGLE(^{0})	35	45	50

GROOVE ANGLE(^{0})	WELDING SPEED (cm/s)
35	0.4
35	0.8
35	1.2
45	0.4
45	0.8
45	1.2
50	0.4
50	0.8
50	1.2

INTRODUCTION TO TAGUCHI TECHNIQUE

- Taguchi defines Quality Level of a product because the Total Loss incurred by means of society due to failure of a product to perform as desired when it deviates from the added goal overall performance degrees.

- This consists of fees associated with negative overall performance, operating prices (which adjustments as a product ages) and any added charges due to dangerous facet consequences of the product in use.

Taguchi Methods

- Help businesses to perform the Quality Fix!

- Quality troubles are because of Noises inside the product or manner machine

- Noise is any unwanted impact that will increase variability

- Conduct tremendous Problem Analyses

- Employ Inter-disciplinary Teams

- Perform Designed Experimental Analyses

- Evaluate Experiments using ANOVA and Signal-to noise strategies

CHAPTER-4

Defining the Taguchi Approach

- Noise Factors Cause Functional Variation
- They Fall Into Three "Classes"

– 1. Outer Noise – Environmental Conditions

– 2. Inner Noise – Lifetime Deterioration

– three. Between Product Noise – Piece To Piece Variation

- The Point Then Is To Produce Processes Or Products The Are ROBUST AGAINST NOISES
- Don't spend the money to put off all noise, build designs (product and manner) that may perform as desired – low variability – within the presence of noise!
- WE SAY:

ROBUSTNESS = HIGH QUALITY

- TO RELIABLY MEET OUR DESIGN GOALS MEANS: DESIGNING QUALITY IN!

TAGUCHI PARAMETER DESIGN FOR TURNING PROCESS

In order to pick out the system parameters affecting the selected device exceptional characteristics of turning, the subsequent process parameters are selected for the present work: reducing pace (A), feed fee (B) and intensity of reduce (C). The selection of parameters of interest and their tiers is based totally on literature evaluate and a few initial experiments conducted.

Selection of Orthogonal Array

The procedure parameters and their values are given in table. It was also decided to look at the 2 – factor interaction outcomes of technique parameters on the selected traits at the same time as turning. These interactions had been taken into consideration among slicing speed and feed charge (AXB), feed price and depth of reduce (BXC), cutting velocity and depth of cut (AXC).

PROCESS PARAMETERS	LE VEL1	LEVEL2	LEVEL3
WELDING SPEED (cm/s)	0.4	0.8	1.20
GROOVE ANGLE(^{0})	35	45	50

Results: Using randomization technique, specimen was turned and cutting forces were measured with the three – dimensional dynamometer. The experimental data for the cutting forces have been reported in Tables. Feed and radial forces being 'lower the better' type of machining quality characteristics, the S/N ratio for this type of response was and is given below

$$S/N\ ratio = -10 \log\left[\frac{1}{n}(y_1^2 + y_2^2 + ... + y_n^2)\right] \quad \ldots (1)$$

Where $y_1, y_2, \ldots, y_n$ are the responses of the machining characteristics for each parameter at different levels

Results

Taguchi technique stresses the significance of reading the response variant the usage of the sign–to–noise (S/N) ratio, resulting in minimization of great characteristic variation due to uncontrollable parameter. The slicing pressure is taken into consideration because the quality feature with the idea of "the larger-the-better". The S/N ratio for the larger-the-higher is:

$$S/N = -10 * \log(\Sigma(Y2)/n))$$

Where n is the range of measurements in a tribulation/row, in this case, n=1 and y is the measured value in a run/row. The S/N ratio values are calculated by means of taking into account above Eqn. With the help of software program Minitab 17.

5.CONCLUSION

The experiment designed by Taguchi method fulfills the desired objective. Fuzzy interference system has been used to find out the ultimate tensile strength .The all possible values of have been calculated by using MINITAB 17.0 software. Analysis of variance (ANOVA) helps to find out the significance level of the each parameter. The optimum value was predicted using MINITAB-17 software.

The welding parameters are Welding speed, and groove angle for TIG welding of work piece steel. In this work, the optimal parameters of welding speed are 0.4cm/s, 0.8 cm/s & 1.2 cm/s, groove angle 35,45 and 50 degrees. Experimental work is conducted by considering the above parameters. Ultimate tensile strength validated experimentally.

The experimental results confirmed the validity of the used Taguchi method for enhancing the welding performance and optimizing the welding parameters in TIG welding at welding speed 1.2 cm/s , and groove angle 35.

6.REFERENCES

1.Arpaci, V. S., 1995, "Buoyant Turbulent Flow Driven by Internal EnergyGeneration," Int. J. Heat Mass Transfer, Vol. 38, pp.

2.761-2770.Asfia, F. J., and Dhir, V. K., 1994, "An 2.Experimental Study ofNarural Convectionin a Volumetrically Heated Spherical Pool With Rigid Wall," ASME Paper 94-WA/H7--26.

3.Cheung, F. B., i980a, "HeatSource-Driven Thermal Convection at Arbirrary Prandtl Numbers," J. Fluid Mech., YoL 97, pp. 743-758.

4. Cheung, F. B., 1980b, "The Boundry Layer Behavior in Transient Turbulent Thermal Convection Flow," ASME JoURNAL oF HEAr TRANSFERV, ol. 102, pp.373-375.

5.Cheung, F. B., 1978, "Turbulent Thermal Convection in a Horizontal FluidLayer With Time Dependent Volumetric Energy Sources." AIAA/ASME Thermo physics and Heat Transfer Conf.,78-HT-6, Palo Alto.

6. Cheung, F.8., 1977, "Natural Convection in a Volumerrically Heared FluidLayer at High Rayleigh Numbers," Int. J. Heat Mass Transfer, Vol. 20, pp.499-506.

Modern Technology in Solar Energy Generation

Nimmala Mahesh[1],P. Manikya rao [2]
Dr. M.G.R Educational and Research Institute, [1,] Mallareddy college of Engineering[2]

Abstract:**Solar energy is playing a pivotal role in compensating the electrical energy as there is short fall in this engergy due to more demand and decline trends of conventional source of energies exhaustion of fuels like coal, petroleum, natural gases and constant of environmental and climatic changes to cope up this photovoltaic installation is being done in an electrical system to compensate and enhance the energy. an photovoltaic installation in an electrical system is made from the assembly of various photovoltaic units that uses solar energy to produce the electricity in a cheaper way from sun power. Till now the use and scope of solar energy is limited and has not reached upto masses Moreover the efficiency of the system is also low due to which the output is not sufficient as compared to input as in some installed case of solar panel it has been observed that efficiency is not more that 27%. To make it versatile and more useful for the masses newer trends and innovations will help. These have discussed in this paper.**

Keywords:**Solar panels, Efficiency, Namadd, renewable engergy resources, distributed generation**

I. INTRODUCTION

Now-a-days distributed generation (DG) is not a new concept. Without this the rectification of energy crises is not possible [1]. In most of the countries the electrical power demand is more than the electrical power generated. On the other hand there is a serious decline in the availability of natural resources, fuels, coal and gases etc. The generation of hydro power plant also varies due to variation in inflow of water from catchment area. When the capacity of hydro power plant decreases the power shortage arises. The solar power plant may be installed in such a fashion that these may work in unison for example when draught is more showering of sun is more. In this way shortage of power will be compensated by the energy governed by the solar power plant. Moreover this installation is to be done in such a way that solar panel will cover the rivers or reservoir reducing the evaporation which enhancing the capacity of dam. On the other hand the solar panel covering this area will generate electrical power which will the enhance the power generation of the system. Moreover by using some newer technologies this power generated may be integrated with the power grid to enhance the capacity of grid.

The development and uses of solar energy at large scale is not only reasonable method of energy resources utilization in the future but also effective frame to improving energy resource crises economically. There is different solar resource in different area, seasons, and weather conditions and so on because of so many influencing factors [2]. As this of kind of renewable energy is available in abundance in nature. The research and application of solar energy will be used to deal with alternative energy [3] – [4]. The advantages of renewable energy sources are enormous as they are free from gas emissions from few conventional energy resources which have impact on the global warming. If this generation of solar energy expedited rigorously can meet the most of the energy demand of the world. Use of Renewable energy will prove as panacea for solving the climatic and environmental problems as every sector of society is keen to solve these problems on the health ground problems. Currently, renewable energy sources install 15 percent to 20 percent of the world's total energy demand [5]. The solar energy is considered as the most promising and important renewable. It is envisaged that solar energy power plants would meet all human needs and would eventually replace the conventional power plants [6].

II. TYPES OF DISTRIBUTED GENERATION

There are different types of distributed generation according to the constructional and technical points of view as shown in Figure (1) [1].

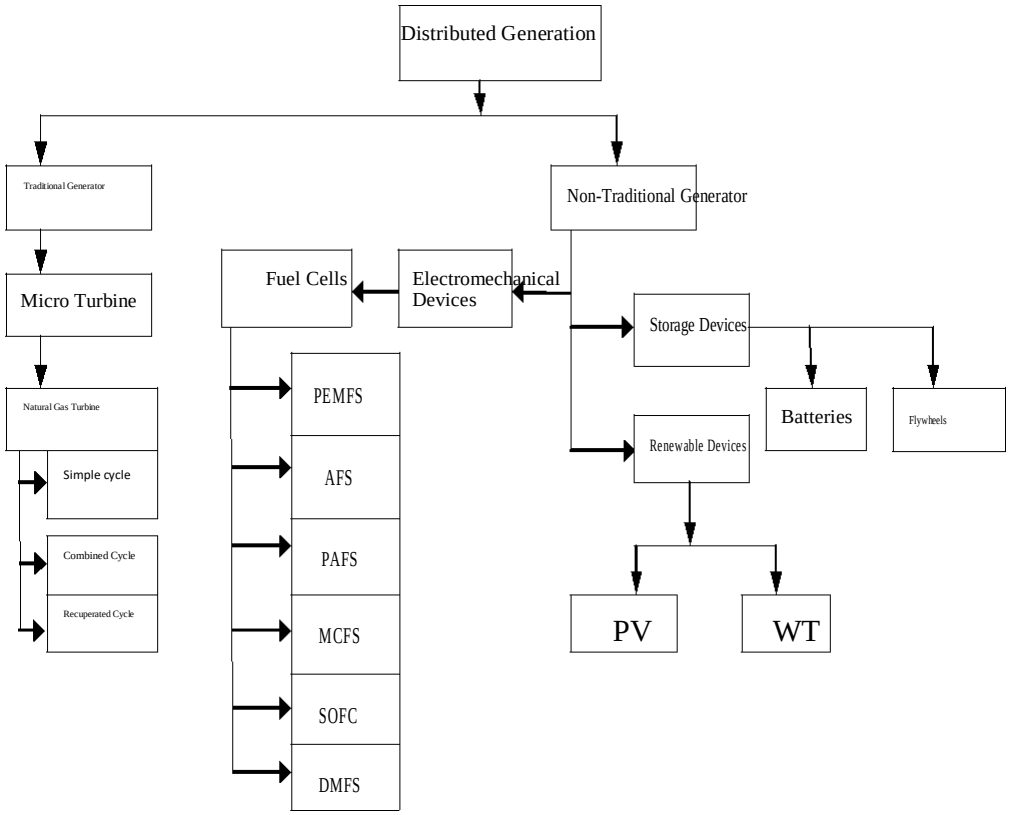


Figure 1 Types of Distributed Generation

III. The Principle Of Energy Storage

The circulation medium was heated by synchronous tracking and non-tracking solar collector and injected into the heat exchanger which was set in concrete pile with the help of circulation pump. After heat exchanging between the heat exchanger and concrete pile, circulation medium was pumped to synchronous tracking solar collector, so formed circulations.And then, the solar energy which was gathered by synchronous tracking and non-tracking solar collector was stored constantly in the underground concrete energy storage pile [7]. The circulating principle is shown in Figure(2). The solar nstal are used to generate steam which drives the steam turbines coupled with alternator as per figure (3).

IV. Solar Energy Policies Of India

The Government of India has increased its focus on developing alternativeresources of energies especially Solar Energy under the policies related to energydevelopment. The solar energy is available in abundance and almost free of cost as it is available from nature. Due to rapid economic expansion India is one of the

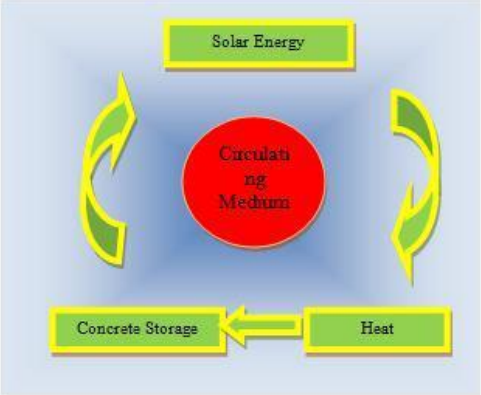


Figure 2 the circulating principle of solar energy

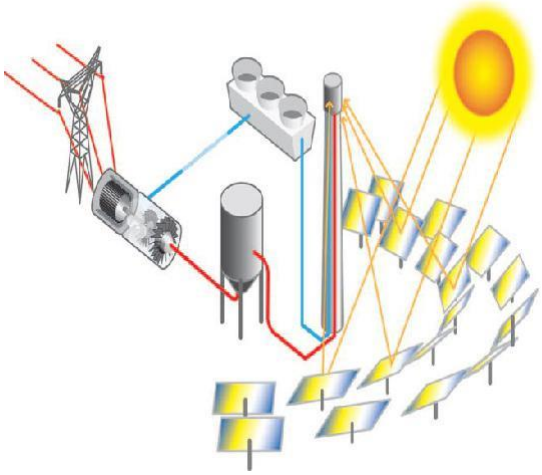


Figure 3 System of Power generation from solar energy

most growing markets and expected to be second largest energy contributor in energy market in the world by 2035. Due to limited domestic fossil fuels reserve, the India has strong planning to expand the renewable energy sources for power sector.

1. To supply the electricity to all the areas included the rural areas as mandated in section 6 of electricity act. Both the Central and State Government will jointly nstalled to achieve this objective at the earliest. Rural Electrification will be done for securing electricity access to the entire household in rural sector. Most of this requirement will be fulfilled by use of renewable energy sources.
2. Reliable rural electrification would be done either through conventional or non conventional methods of electricity whichever is more suitable and economical. Non conventional sources of energy especially Solar can be utilized even where Grid connectivity exists [8].
3. Particular attention is to be given to Dalit Bastis, Tribal areas and other weaker sections of the society the other newer resources.
4. Rural Electricity Corporation of India (REC) is the nodal agency at central govt. Level to implement these programs of electrification in rural areas. The REC will nstal all the goals set up by the National Common Minimum Programme ensuring timely implementation [9].
5. Responsibility of operation and maintenance & cost recovery could be discharged through appropriate arrangement with Panchayats, Local Authorities, BDO, and NGO etc [10].

6. This Great task of Rural Electrification requires cooperative efforts of all agencies like Govt. Of India, State Government and community education cell in rural areas.
7. The Electricity act 2003 has provision of restructuring the electricity industry which unbundled the vertically integrated electricity supply in each state. Now generation, transmission and distribution companies have been formed by the Regulatory Commission of state electricity board. Regulatory Commission will also specify the minimum percentage of electricity that each distribution utility must get from renewable energy sources [11].

V. Newer Methods Which Will Enhance the Use of Solar Energies

Day by day new trends and innovations are being developed throughout the world in R&D centers, automobile sectors and domestic use in institutions, hostels to reduce the energy wastage and to generate the power by solar devices. Many of them are explained given below:

1. Solar cells of higher efficiency have been developed having conversion efficiency more than 37% as compared to the previous solar cells having efficiency of 27 % made of two materials. Tata power is going to install the solar panels having 35% efficiency [12]. In These cells three photo absorption layers are stacked together. This has been developed by stacking Indium, Gallium and arsenide as the bottom layers. These cells have capability of absorbing the light from various wavelengths available in sunlight and convert into electrical energy. Through optimal process the active area has been increased. This breakthrough in technology has been done by new energy and industrial technology development organization.

1. The conversion efficiency of solar panels/plates is increased by newer devices of cleaning these panels. The device makes use of automated "dry-sweep'' to push dust and dirt away from the surface of these devices. In south Arabian language the device is known as nsta which is very rugged and have low maintenance cost. The device is powered by the lithium ion batteries. These batteries are charged by the array itself and have high efficiency. The device has moving parts. It is very interesting that this act like a robotic arm and automated work with scheduling. The device can jump the obstacle between the panels[13].

2. Throughout the world research and development is done to procure more and more energy from various devices and technologies. Under this concept Japan developed a fabric which is known as a solar cell fabric capable of harnessing the energy from sunlight while you are moving by wearing this fabric made cloth. This fabric is made from wafer thin solar cells woven in a stylish way. The electricity generated will be capable to charge the mobile and other portable electronic gadgets. The thread used will be stronger and which increases the life or durability of the fabric cloth. The same idea can be embedded/used in the blind makers and certain type of curtain will also generate power, when sun rayeson these. Various companies developing this type of fabric in association with solar cell maker. This will help the men to recharge these small gadgets while in sun [14].

3. A new trend of solar panel roofs have come in which most of the buildings the roof is covered with solar panel. In some advanced countries like china the roof of the max. Of the homes is made of solar panel by using aaluminum or strong alloy to support the weight of panels. In remote area where grid supply is not viable these panels generate power for themselves and supply electricity to the neighbours also who cannot afford the cost of installation. This will help nearby masses and community in that region where distribution of power is not feasible by other ways & transmission may not be possible due to heavy expenditure.

4. In some countries in urban area some hobbyist of solar energy generate electrical power this system for sufficient for their requirement and surplus generated power is supplied to the grid empowering the national grid. The solar panel should be installed on the vehicles where it is possible so that charging of batteries and other devices in the vehicle may be done with the help of solar energy. Whenever the solar rays fall on the panel this will improve the electrical efficiency of the vehicles. This type of experimentation and uses are already being done in Japan and in other advanced countries where conversion technologies from solar to electrical are being used frequently and sufficiently. In India also solar panels have been installed in metro railway service.

5. As we are aware of that electrical power demand is increasing and viewing the climatic concerns it is desired that renewable energy sources especially solar may be integrated to the utility grid. By using better flexibility in integration through power electronics. Harmonics can be reduced and the reactive power can be balanced.

6. In these days most of the power industries switchover side by side starting manufacturing of the solar inverters of high capacity. These may be utilized to get emergent power if not continues at remote locations where there is no grid supply.
7. CSP systems technology is used for power generation in the system large, flat, sunlight mirrors known as heliostats receive sun light at the top of the tower. A fluid for heat transfer is used to generation the steam which is used for production of electrical power .In some countries the capacity of these plant as high as up to 200 MW. These power tower are very popular in these days because of solar to electrical conversion efficiency is high [15].

VI. Distributed Solar Energy Generation

In [16] the authors has explained the use of Distributed solar photovoltaic (PV) systems is producing electricity onsite, so reducing the requirement to build up new transmission line and also avoiding line losses. Distributed generation also offer significant benefits to the consumers while providing resiliency to an electric grid that is based on the traditional and centralized model. These systems are used in applications ranging from small commercial to residential and for industrial use. Though this market is still primarily driven by government incentives, distributed solar PV will continue its steady march in future. Due to reduced market activity in Italy and Germany, global distributed solar photovoltaic market contracted slightly in 2012, However, growth in the United States, China, Japan, and other countries continued, driven by solar PV module price reductions, the growth of third-party financing models, and feed-in tariffs. Navigant Research forecasts that, from 2013 to 2018, 220 GW of distributed solar PV will be installed worldwide, representing $540.3 billion in revenue.

VII. Conclusion

Due to decline aviability of natural's fuels and viewing environmental changes causes due to conventional method of generation, the use ogf solar energy is becoming popular and urgency of the day. This will create healthy environment for the humain beings which are suffering from the various hazards due to pollution from the nstalled contents. Moreover the power generation due to hydro power plant is not also regular due to irrugalar flow of water from the catchment area. So it is concluded that solar power plant may be nstalled in such a way so these may work in unison with hydro and other methods of generation to enhance the clean and green energy.

References

[1] W El-Khattam and M.M.A Salama, "Distributed generation technologies, definitions and benefits", Electric Power Systems Research, vol. 71, Issue 2, p.p.,119–128, October 2004.
[2] Mingzhi Zhao, Zhizhang Liu and Mingjun Yu, "Data acquisition and analyzing of solar energy resource", Information and Automation (ICIA), Harbin, p.p., 2205-2208, 20-23 June 2010.
[3] Zhao Qingbo, Shan Baoguo, Situation and Outlook of Energy Demand in the World, Energy of China, 2002(02):34-36.
[4] Zehner and Ozzie, "Green Illusions", Lincoln and London: University of Nebraska Press. Pp., 331–42, 2012
[5] G.M. Shafiullah, A.M.T. Oo, D. Jarvis, .B.M.S. Ali, P. Wolfs, "Prospects of Solar Energy in Australia", Electrical and Computer Engineering (ICECE), 2010 International Conference, Dhaka p.p., 350 – 353, 18-20 Dec. 2010.
[6] S. Rao and B.B.Parulekar, "Energy Technology Non-Conventional, Renewable & Conventional", Khanna Publication, 3rd, 2012.
[7] Yan Zhao, Dajun Zhao, Yirong Yao, Xin Fang ,Shousheng Li, "Development of Solar Energy Underground Seasonal Storage Device and Its Parameters Measuring System", Information Engineering (ICIE), 2010 WASE International Conference, Beidaihe, Hebei, p.p., 231 – 234, 14-15 Aug. 2010.
[8] G. D. Kamalapur and R. Y. Udaykumar, "Rural Electrification in the Changing Paradigm of Power Sector Reforms in India", International Journal of Electrical and Computer Engineering, p.p., 147-154, vol. 2, No.2, April 2012.
[9] N. Sasidhar, "Electricity online trading in India", Rural Electricity Corporation of India, 2012.
[10] Giuseppe Buglione, Guido Cervigni, Eileen Fumagalli, Elena Fumagalli and Clara Poletti, "Integrating European Electricity Markets", Center for Research on energy and environmental economics policy, Report 2, Oct. 2009.
[11] MadhuKhanna, KusumMundra, AmanUllah, "Parametric and Semi-Parametric Estimation of the Effect of Firm Attributes on Efficiency: Electricity Generating Industry in India", Journal of International Trade and Economic Development. Pp. 419–436.vol. 8, No. 4, Sept. 2011.
[12] Electronics for You, "Most efficient Solar Cell", Technology news, January, 2013.

[13] GeorgeEitelhube, “Seed Fund Winners - NOMADD: The No-water Mechanical Automated Dusting Device”, King Abdullah University of Science and Technology (KAUST), energy and sustainability are primary research drivers, 2012.
[14] John Badding, “Flexible Silicon Solar-Cell Fabrics May Soon Become Possible”, journal Advanced Materials,Penn State University, 6 December 2012.
[15] DolfGielen, “Concentrating Solar Power technology”, Energy Technology Systems Analysis Programme, International Renewable Energy Agency, January 2013.
[16] G. D. Rai, “Non-Conventional Energy Sources”, Khanna Publication, 4th edition, 2012.

DESIGN AND STUDY OF A MULTI PLATE CLUTCH

B DILEEPKUMAR
M.Tech (cad/cam)
Malla Reddy college of engineering, maisammaguda
(affiliated to Jntuh)
Hyderabad,Telangana.

ABSTRACT:

Clutch is a mechanism which transfers the rotary motion of one shaft to the other shaft when desired. In automobiles friction clutches are widely used in power transmission applications. To transmit maximum torque in friction clutches selection of the friction material is one of the important tasks.

In this thesis, the general introduction to the arrangement, design and some basic concept of multi plate wet type clutch. Fluid plays an important role in this type of clutch so some of their properties are discussed. Some losses due to design parameters are also discussed. To meet the requirements of low fuel consumption, good driving performance and manufacturing feasibility. This paper will provide a design overview of the transmission architecture, main characteristics, key subsystems and control strategies. This gives a better understanding about working principle of clutch, material used for making the clutch plates. Effect of design consideration can be further studied during its application in various conditions.

In this paper, we design a multi plate clutch by using empirical formulas. A 2D drawing is drafted for multi plate clutch from the calculations & a 3D model is created in the modeling Software CREO for Automobile Applications. Analysis done in ANSYS with different materials.

Static analysis to determine the deformation, stress and strain for the single plate clutch. Modal analysis is to determine the deformations with respect to frequencies.

Keywords: 2d 3d design, materials, stress, strain, friction, power transmission.

CHAPTER:1

INTRODUCTION:

The clutch is a mechanical device, which is used to connect or disconnects the source of power from the remaining parts of the power transmission system at the will of operator. The clutch can connect or disconnect the driving shaft and driven shaft. An automotive clutch can permit the engine to run without driving the car. This is desirable when the engine is to be started or stopped, or when the gears to be shifted. Clutch is a mechanism for transmitting rotation, which can be engaged and disengaged. The clutch connects the two shafts so that they can either be locked together and spin at the same speed (engaged), or be decoupled and spin at different speeds (disengaged). Depending on the orientation, speeds, material, torque produced and finally the use of the whole device, different kinds of clutches are used. The clutch in itself is a mechanism, which employs different configurations. The friction clutch is an important component of any automotive machine. It is a link between engine and transmission system which conducts power, in form of torque, from engine to the gear

assembly. When vehicle is started from standstill clutch is engaged to transfer torque to the transmission; and when vehicle is in motion clutch is first disengaged of the drive to allow for gear selection and then again engaged smoothly to power the vehicle. Generally there are two types of clutches based on type of contact Positive clutch− Friction clutch− Single plate comes under the category of friction clutch. Desirable properties for friction materials for clutches: The two materials in contact must have a high− coefficient of friction.

INTERLOCKING PARTS CLUTCHES:

This type of clutch has protruding circular edge and a hole for them that engages and disengages during operation. This type is less effective since human foot or hand power on clutching reaches about 10 KN or 1,000 kg.

FRICTION CLUTCHES :

A friction clutch The vast majority of clutches ultimately rely on frictional forces for their operation. The purpose of friction clutches is to connect a moving member to another that is moving at a different speed or stationary, often to synchronize the speeds, and/or to transmit power. Usually, as little slippage (difference in speeds) as possible between the two members is desired.

CHAPTER:2

APPLICATIONS:

MACHINERY:

This type of clutch is used in some lawnmowers, copy machines, and conveyor drives. Other applications include packaging machinery, printing machinery, food processing machinery, and factory automation.

AUTOMOBILES:

When the electromagnetic clutch is used in automobiles, there may be a clutch release switch inside the gear lever. The driver operates the switch by holding the gear lever to change the gear, thus cutting off current to the electromagnet and disengaging the clutch. With this mechanism, there is no need to depress the clutch pedal. Alternatively, the switch may be replaced by a touch sensor or proximity sensor which senses the presence of the hand near the lever and cuts off the current. The advantages of using this type of clutch for automobiles are that complicated linkages are not required to actuate the clutch, and the driver needs to apply a considerably reduced force to operate the clutch. It is a type of semi-automatic transmission.

LOCOMOTIVES :

Electromagnetic clutches have been used on diesel locomotives, e.g. by Hohenzollern Locomotive Works.

CHAPTER:3

OTHER TYPES OF ELECTROMAGNETIC CLUTCHES:

MULTIPLE DISK CLUTCHES:

Introduction – Multiple disk clutches are used to deliver extremely high torque in a relatively small space. These clutches can be used dry or wet (oil bath). Running the clutches in an oil bath also greatly increases the heat dissipation capability, which makes them ideally suited for multiple speed gear boxes and machine tool applications.How it works – Multiple disk clutches operate via

an electrical actuation but transmit torque mechanically.

ELECTROMAGNETIC TOOTH CLUTCHES:

Of all the electromagnetic clutches, the tooth clutches provide the greatest amount of torque in the smallest overall size. Because torque is transmitted without any slippage, clutches are ideal for multi stage machines where timing is critical such as multi-stage printing presses. Sometimes, exact timing needs to be kept, so tooth clutches can be made with a single position option which means that they will only engage at a specific degree mark. They can be used in dry or wet (oil bath) applications, so they are very well suited for gearbox type drives.

ELECTROMAGNETIC PARTICLE CLUTCHES:

Introduction Magnetic particle clutches are unique in their design, from other electro-mechanical clutches because of the wide operating torque range available. Like a standard, single face clutch, torque to voltage is almost linear. However, in a magnetic particle clutch torque can be controlled very accurately. This makes these units ideally suited for tension control applications, such as wire winding, foil, film, and tape tension control. Because of their fast response, they can also be used in high cycle applications, such as card readers, sorting machines, and labeling equipment.

AUTOMOBILE POWERTRAIN:

This plastic pilot shaft guide tool is used to align the clutch disk as the spring-loaded pressure plate is installed. The transmission's drive splines and pilot shaft have a complementary shape. A number of such devices fit various makes and models of drivetrains.

MOTORCYCLES:

A BASKET CLUTCH:

Motorcycles typically employ a wet clutch with the clutch riding in the same oil as the transmission. These clutches are usually made up of a stack of alternating plain steel and friction plates. Some plates have lugs on their inner diameters that lock them to the engine crankshaft. Other plates have lugs on their outer diameters that lock them to a basket that turns the transmission input shaft. A set of coil springs or a diaphragm spring plate force the plates together when the clutch is engaged.

On motorcycles the clutch is operated by a hand lever on the left handlebar. No pressure on the lever means that the clutch plates are engaged (driving), while pulling the lever back towards the rider disengages the clutch plates through cable or hydraulic actuation, allowing the rider to shift gears or coast. Racing motorcycles often use slipper clutches to eliminate the effects of engine braking, which, being applied only to the rear wheel, can cause instability.

CHAPTER:4

OTHER CLUTCHES AND APPLICATIONS:

Belt clutch: Used on agricultural equipment, lawn mowers, tillers, and snow blowers. Engine power is transmitted via a set of belts that are slack when the engine is idling, but an idler pulley can tighten the belts to increase friction between the belts and pulleys.

Dog clutch: Utilized in automobile manual transmissions mentioned above. Positive

engagement, non-slip. Typically used where slipping is not acceptable and space is limited. Partial engagement under any significant load can be destructive.

Hydraulic clutch: The driving and driven members are not in physical contact; coupling is hydrodynamic.

SPECIALTY CLUTCHES AND APPLICATIONS:

SINGLE-REVOLUTION CLUTCH:

Single-revolution clutches were developed in the 19th century to power machinery such as shears or presses where a single pull of the operating lever or (later) press of a button would trip the mechanism, engaging the clutch between the power source and the machine's crankshaft for exactly one revolution before disengaging the clutch. When the clutch is disengaged and the driven member is stationary. Early designs were typically dog clutches with a cam on the driven member used to disengage the dogs at the appropriate point.

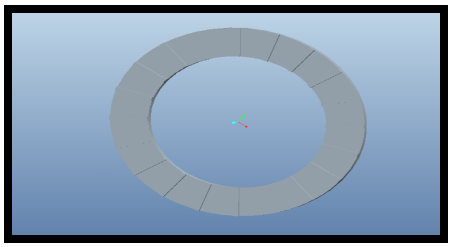
CASCADED-PAWL SINGLE-REVOLUTION CLUTCHES: Cascaded-pawl single-revolution clutch driving the cam cluster in a Teletype Model 33 that performs fully mechanical conversion of incoming asynchronous serial data to parallel form. The clutch drum, lower left, has been removed to expose the pawls and trip projections.

These superseded wrap-spring single-revolution clutches in page printers, such as teleprinters, including the Teletype Model 28 and its successors, using the same design principles.

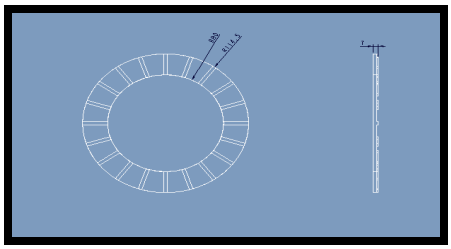
KICKBACK CLUTCH-BRAKES:

These mechanisms were found in some types of synchronous-motor-driven electric clocks. Many different types of synchronous clock motors were used, including the pre-World War II Hammond manual-start clocks. Some types of self-starting synchronous motors always started when power was applied, but in detail, their behaviour was chaotic and they were equally likely to start rotating in the wrong direction. Coupled to the rotor by one (or possibly two) stages of reduction gearing was a wrap-spring clutch-brake. The spring did not rotate. One end was fixed; the other was free. It rode freely but closely on the rotating member, part of the clock's gear train. The clutch-brake locked up when rotated backwards, but also had some spring action. The inertia of the rotor going backwards engaged the clutch and wound the spring. As it unwound, it restarted the motor in the correct direction. Some designs had no explicit spring as such—but were simply compliant mechanisms. The mechanism was lubricated and wear did not present a problem.

3D MODEL:



2D MODEL:



STATIC ANALYSIS OF SINGLE PLATE CLUTCH:

Materials used

Steel:

Young's modulus = 205000mpa

Poisson's ratio = 0.3

Density = 7850kg/mm^3

Cast iron:

Young's modulus = 110000 mpa
Poisson's ratio = 0.28

Density = 7200

Copper :

Young's modulus = 101000mpa

Poisson's ratio = 0.32

Density = 6800

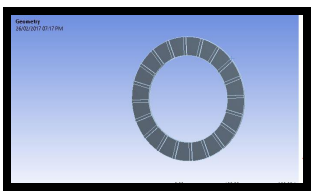
Save Pro-E Model as .iges format

→→Ansys → Workbench→ Select analysis system → static structural → double click

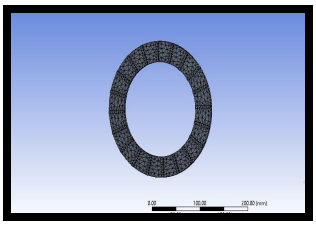
→→Select geometry → right click → import geometry → select browse →open part → ok

→→ Select mesh on work bench → right click →edit

Double click on geometry → select MSBR → edit material →



Select mesh on left side part tree → right click → generate mesh →



Select static structural right click → insert → select rotational velocity and fixed support →

5 . RESULTS TABLES:

STATIC ANALYSIS RESULTS

MATERIAL	DEFORMATION(mm)	STRESS (N/MM^2)	STRAIN
STEEL	1.0788E-5	0.3023	1.6012E-6
CAST IRON	1.92E-5	0.31251	2.84E-6
COPPER	1.54E-5	0.32894	2.452E-6

FATIGUE ANALYSIS RESULTS

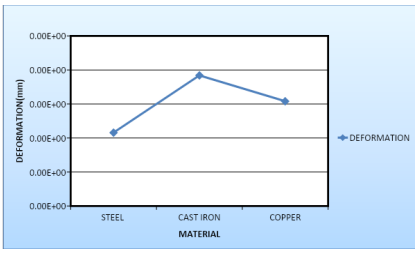
MATERIAL	LIFE	DAMAGE	SAFETY FACTOR
STEEL	1E10	0.1	4.3094
CAST IRON	1E10	0.1	4.4159
COPPER	1E10	0.1	4.1953

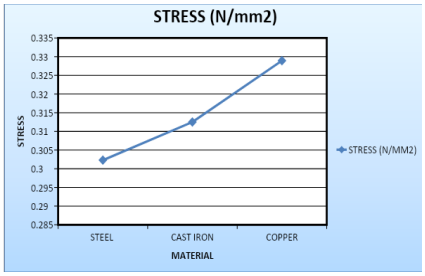
MODAL ANALYSIS RESULTS

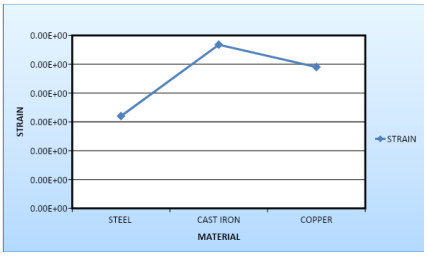
MATERIAL	frequency	deformation 1	frequency	Deformation2	frequency	Deformation 3
STEEL	1.3133E+05	589.73	1.314E+05	655.2	1.31E	626.29

					+ 05	
CAST IRON	1.0233E +005	616.95	1.023E+05	691.08	1.024E+05	659.93
COPPER	93596	569.86	93665	623.97	93696	597.78

GRAPHS







6 . CONCLUSION:

Static Structural, modal and fatigue analysis is done for multi plate clutch using the properties of the three materials. Outer material is steel and Materials used for friction plate are varied Cast iron, Cork and Asbestos. Comparison is done for above materials to validate better friction material for multi plate clutch. By observing static analysis results, the stress values for all materials are less than that of their respective yield stress values. The deformation and stress values are less for Cast Iron and Asbestos is used. By observing modal analysis results, the deformation is less for Cast Iron but the frequencies are less when Cork is used. Since the frequencies are lesser, the vibrations in the clutch will be reduced when cork is used. By observing fatigue analysis results, the life is more for Cast Iron but the damage is more for Cork and Asbestos. The clutch will be failed if the applied load is multiplied with the damage value. Since the damage value is more for Cork and Asbestos, the clutch when both the materials are used will fail at very larger loads. So it can be concluded that using Cork for friction plate is better.

6 . REFERENCES:

[1] Mr.P.V.Narasimha Rao.L, Ch.Chandrarao ―Static and Dynamic Analysis of Clutch Plate With Crack‖, IJRMET, Volume 4, Issue Spl - 1, Nov 2013

[2] Mr.Ganesh Raut, Anil Manjare, P Bhaskar ―Analysis of Multidisc Clutch Using FEA‖ IJETT, Volume 6, Dec 2013

[3] Prof.P.Baskar, Anil Jadhav, Gauri Salvi, Santosh Ukamnal, ―Static Structural Analysis of Multi plate Clutch with Different Friction Materials‖ IJERT, Volume 2, Issue 11, November 2013

[4] Dr. J. P. Modak, Shrikant V.Bhoyar, Dr. G. D. Mehta, ―Dynamic Analysis Of Single Plate Friction Clutch‖, IJERT, Vol. 2, Issue 7, July – 2013

[5] Mr.Prashil M. Mhaiskar, Nitin D. Bhusale, Mayur D. Pastapure ―Vibration Analysis of Dry Friction Clutch Disc by Using Finite Element Method‖, IJERT, Volume 3, Issue 1, January – 2014

[6] Mr.Rajesh Purohit, Pooja Khitoliya and Dinesh Kumar Koli, ―Design and Finite Element Analysis of an Automotive Clutch Assembly‖, ScienceDirect, Procedia Materials Science 6, 2014

[7] Mr.Monarch K. Warambhe, Gautam R. Jodh, Mamta G. Pawar, ―Design and Analysis of Clutch Using Sintered Iron as a Friction Material‖ ,IJITEE, Volume-3, Issue-7, December 2013

[8] Mr. P.Naga Karna, Tippa Bhimasankara Rao, ―Analysis of Friction Clutch plate using FEA‖ ,IJERD, Volume 6, Issue 2, March 2013

[9] Mr. B. Nivas, M. Nithiyanandam, S. Tharaknath, A. AvinashKumar, ―Design and Analysis of Clutch Plate Using Steel Material [En –Gjs-400 -15steel]‖, IOSR-JDMS, Volume 13, Issue 5, May. 2014

Design Modification of Heat Exchanger Using CFD Tools

[1]Prakash D Chavan, [2]Kandy Sandya

[1]Assistant Professor, Mechanical Department, Malla Reddy college of engineering, Maisammaguda, secunderabad, India.

[2]Assistant Professor, Mechanical Department, Malla Reddy college of engineering, Maisammaguda, secunderabad, India.

Abstract

Heat exchangers are used in the thermal system to maintain the temperature of the working liquid. Among the various types of heat exchanger, the shell and tube heat exchanger is the most commonly used heat exchanger based on its simple design and performance aspects. Even though these shell and tube heat exchangers operates at its designed point, it can be even effectively designed to achieve better heat transfer rate.

Since most of the shell and tube heat exchangers are designed based on the traditional design concepts, in this paper, we have planned to modify the design of the heat exchanger. For this purpose, we have selected a reference heat exchanger with its practical performance results. Using the design data from the existing heat exchanger, the CAD model was generated using solid works software and it was analysed using the CFD software under the actual operating conditions.

Then, the design modifications was carried out in the inner tubes and baffles accordingly. Initially we have changed the design of baffles and based on the analysis results, the cross section of the tube was modified. These modified design was analysed using the CFD tools under the same operating conditions and the results was compared with the actual and modified designs.

By this, we will be justifying the application of CFD tools in the design of heat exchangers to predict its performance in the early design stage. Due to this, the time and money invested on the man and equipments will be reduced to the industry for developing an efficient heat exchanger.

Keywords: Heat exchangers – CFD tools – performance analysis – design optimization.

INTRODUCTION

Heat exchangers are widely used in the thermal systems where there are requirement of maintaining the system temperature in order to get the quality product. The quality of the final product is directly associated with the efficiency and effectiveness of the heat exchanger. The efficiency of the heat exchanger is based on the quantitative values of the inlet and exit temperature difference.

The efficiency of this type of heat exchanger is determined by the amount of heat transferred from hot to cold mediums and in vice versa. Also in some peculiar cases, for the customized purposes, the flow inside the heat exchangers are changes from counter flow to parallel flow. This is purely based on the application where it is been used. In most of the cases, the type of fluid flowing in the shell and in the tubes are also changed to get better efficiency.

Based on the above, this project is based on the performance study of a shell and tube heat exchanger using a CFD tool. This study is carried out with respect to the available practical data. The practical data are taken from the reference paper and the design of the heat exchanger is also taken from the reference paper. The reference paper contains the practical experimental data. This study is to justify the application of CFD tools in the design of the heat exchanger system. The analysis will be carried under varied conditions as per the reference paper.

The first phase of this project deals with the study and analysis of the heat exchangers along with the selection of reference paper continued by the modelling of the heat exchanger system. The second phase of this project will be contained with the CFD analysis of the system using the solid works flow simulation tools. Then the experimental and CFD results will be compared to justify the effective application of the CFD tools in thermal system design.

METHODOLOGY

Methodology is the basic requirement for a project, because it defines the proper start and end conditions of the works to be done. Proper planning and execution of the workflow decides the successful completion of the project. The methodology of this project is as follows.

Figure 1: The strategic planning process

Problem Definition

The problem is defined as to model a heat exchanger as per the physical dimensions available in the industry and to conduct the CFD analysis for the existing operating conditions of 55 degree Celsius and 25 degree Celsius with hot water in shell and cold water in tube. This analysis have to be done under parallel flow conditions and verify the CFD results are same as actual experimental results with general limitations.

Also the analysis should also be conducted for various operating conditions like parallel flow and counter flows and varied baffle geometric conditions.

Objective

The objective of this project is to justify the applications of CFD tools in the thermal system design. Also the baffle conditions have to be optimized accordingly to find the best operating conditions.

Geometric Specifications

The geometry specifications are taken from the design data available in the reference journals [1]. The following are the details of the heat exchanger used for analysis.

Figure 2: Real time experimental setup

Shell Details

Outer diameter = 142mm

Inner diameter = 136mm

Length of the HE = 1500mm

No. of baffles = 5

Distance between baffles = 300mm

Baffle opening = 25% (except first and last)

Tube Details

Outer diameter = 23mm

Inner diameter = 20mm

Length = 1200mm

No. of tubes = 9

Material of construction

Shell = Stainless Steel

Tubes = Copper

Baffles = Copper

Modelling

The following are the CAD images of the Heat exchanger modelled using Solid Works and drafter using Auto CAD.

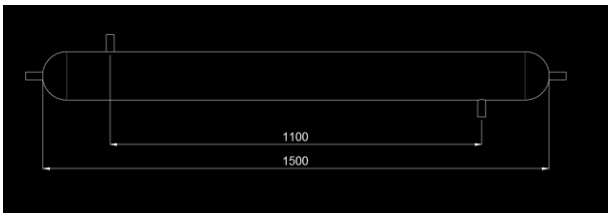


Figure 3: Dimensions of Shell and openings

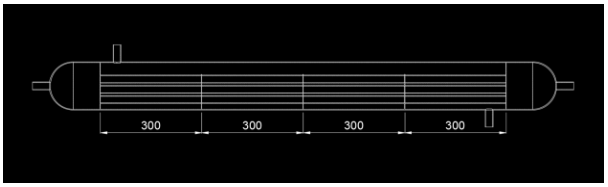


Figure 4: Dimensions of baffle arrangements

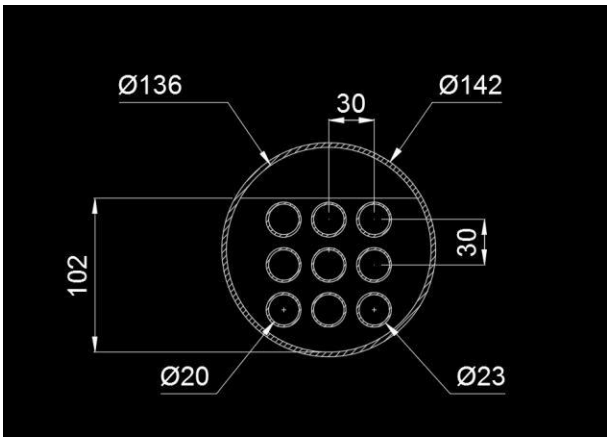


Figure 5: 8 Dimensions of shell, tube and baffle

Figure 6: Front view of the Heat Exchanger

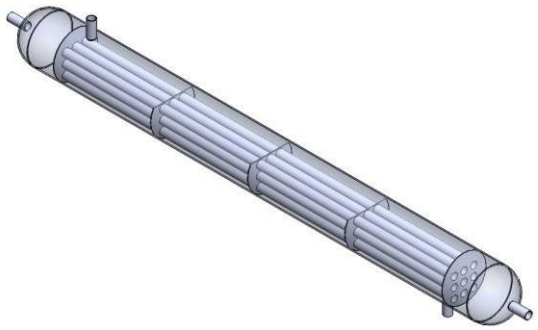


Figure 7: View of tubes and baffles

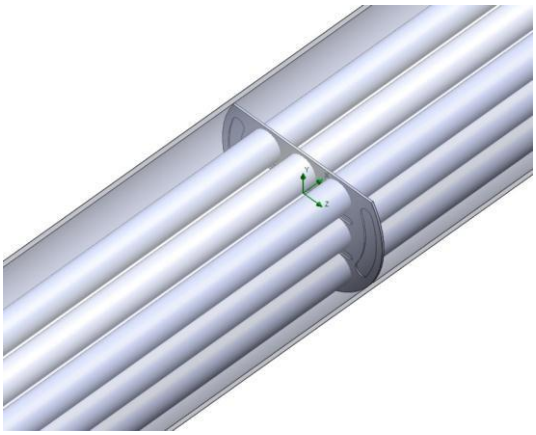


Figure 8: Slots in baffles (slots are of half the thickness)

Analysis

The performance analysis for the heat exchanger was done using Solid Works Fluid Flow Simulation. The analysis was conducted for the following operating conditions.

Case 1: Hot water in Shell and Cold water in Tube – Parallel flow (actual condition)

Case 2: Hot water in Shell and Cold water in Tube – Counter flow

Case 3: Hot water in Tube and Cold water in Shell – Parallel flow

Case 4: Hot water in Tube and Cold water in Shell – Counter flow

Among these analysis, the case 1 is the actual condition in which the system is been operated and tested in the industry. The following are the results of the experimental test.

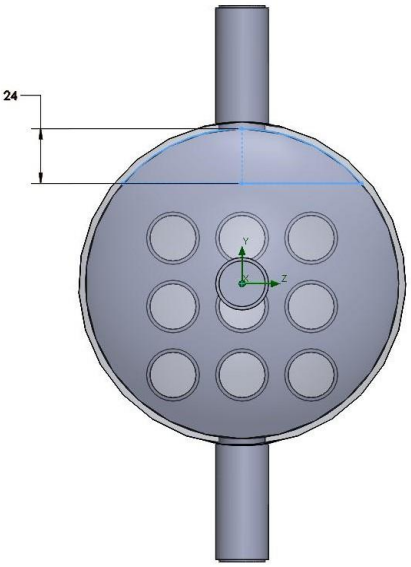


Figure 9: Baffle gap reduced from 34 mm to 24mm

Table 1: Ranked problems of the Russian road transport enterprises and options for solution of these problems

Dated- 11th & 12th of April 2014 (12:00 to 4:30 PM), ambient temperature: 33 (°C)

Serial number	Hot water inlet (°C)	Hot water outlet (°C)	Cold water inlet (°C)	Cold water outlet (°C)	Degree of cooling (°C)	Degree of heating (°C)	Effectiveness of heat exchanger
1.	55.3	47.5	25.5	31.1	7.8	5.6	0.26174
2.	55.4	47.9	24.2	30.3	7.5	6.1	0.24038
3.	55.0	47.5	25.0	30.0	7.5	5.0	0.25
4.	55.3	47.5	25.5	31.1	7.8	5.6	0.26174
5.	55.4	47.9	24.2	30.3	7.5	6.1	0.24038
6.	55.7	48.1	24.7	30.0	7.6	5.3	0.24516
7.	55.3	48.3	25.2	29.3	7.0	4.1	0.23255

Form the above set, the set 3 is taken for CFD analysis and for comparison purpose. The above set is for case 1 analysis. The analysis results of all the four cases will be compared and the best operating conditions will be selected for the further development activities. The following are the proposed development models.

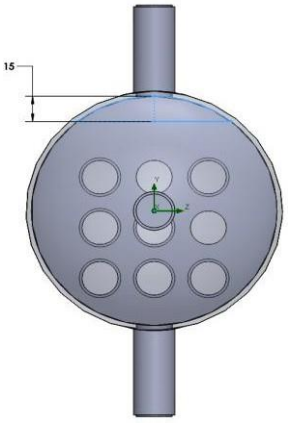


Figure 10: Baffle gap reduced from 34 mm to 15mm

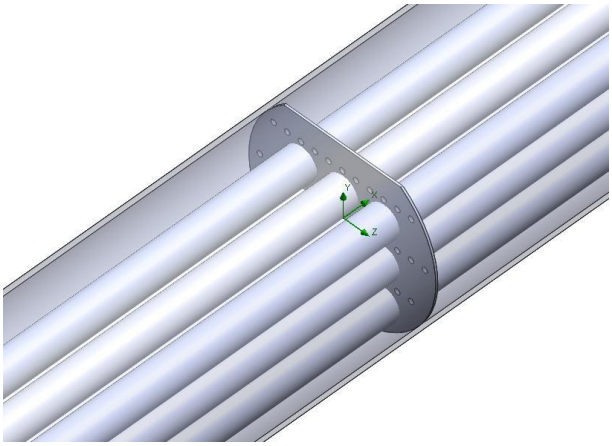


Figure 11: Baffle gap 15mm along with spherical dimples on baffles

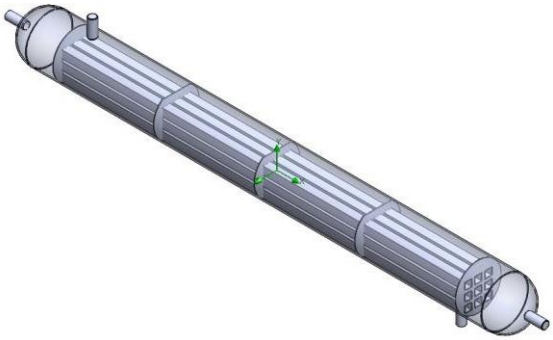


Figure 12: Proposed change in tube shape _ Rectangular shape tubes

This rectangular shaped tube design will be implemented in the design which is having better performance among those which have been mentioned earlier.

The CFD analysis was conducted under the following 3 steps.

- Pre-processing
- Solution
- Post-processing

The model importing and cleaning, meshing, boundary conditions and material property assigning are all done at the pre-processing stage. The solver settings and output settings and simulations are carried at the solution stage. The extraction of results from the saved database in the form of contour plots and tabulated values are done in the post-processing and this post-processing will be explained in the upcoming chapter.

The boundary conditions used for the analysis are as follows.

1. Hot liquid inlet = 55 0C
2. Cold liquid inlet = 25 0C
3. Mass flow rate (Hot) = 0.027 Kg/sec
4. Mass flow rate (Cold) = 0.014 Kg/sec
5. Ambient temperature = 30 0C
6. Hot and cold liquid = Water

Assumptions made are:

1. Flow is laminar and turbulent
2. U = 750 W/m2K (based on data available in net)

RESULTS AND DISCUSSION

The following are the typical CFD contour plot outputs. These outputs are retrieved using the post processing tools. In total 9 cases have been analyzed and its quantitative results are displayed as tabular columns and histogram charts following the below schematic CFD plots.

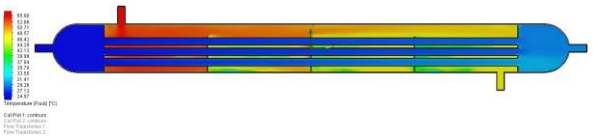


Figure 13: Interior fluid temperature (front plane cut section)

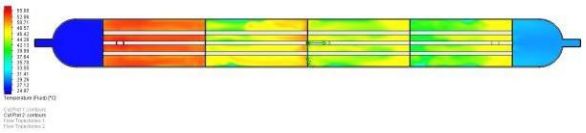


Figure 14: Interior fluid temperature (top plane cut section)

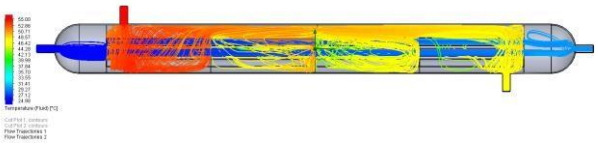


Figure 15: Interior fluid flow pattern (flow trajectories)

Table 2: Results of Case 1 to Case 4 (actual model with varied operating conditions)

CONDITIONS	HOT OUTLET [°C]	COLD OUTLET [°C]
CASE 1	46.845	29.393
CASE 2	46.689	29.594
CASE 3	48.522	28.913
CASE 4	48.473	29.201

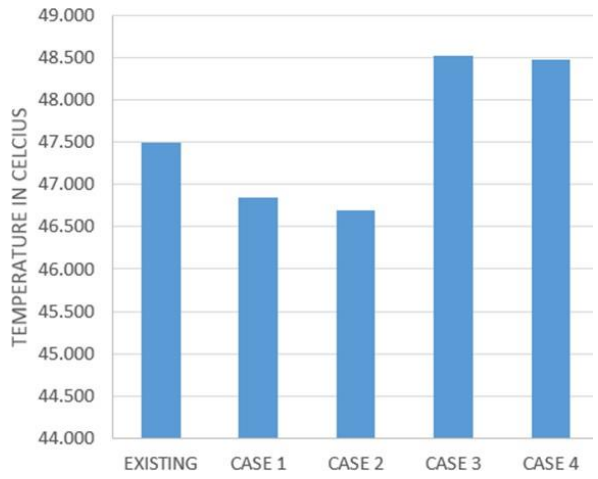


Figure 16: Hot water outlet temperatures – case 1-4

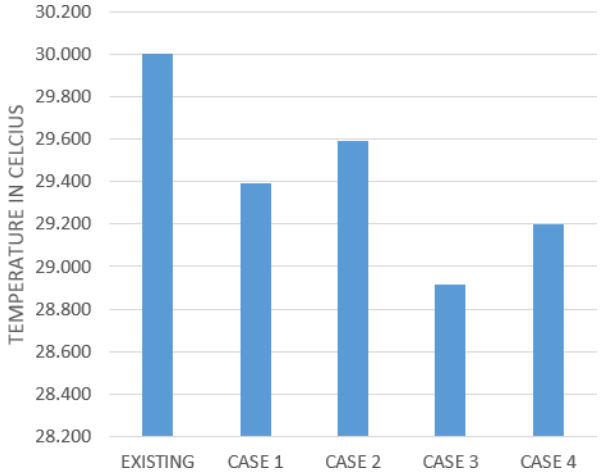


Figure 17: Cold water outlet temperatures case 1-4

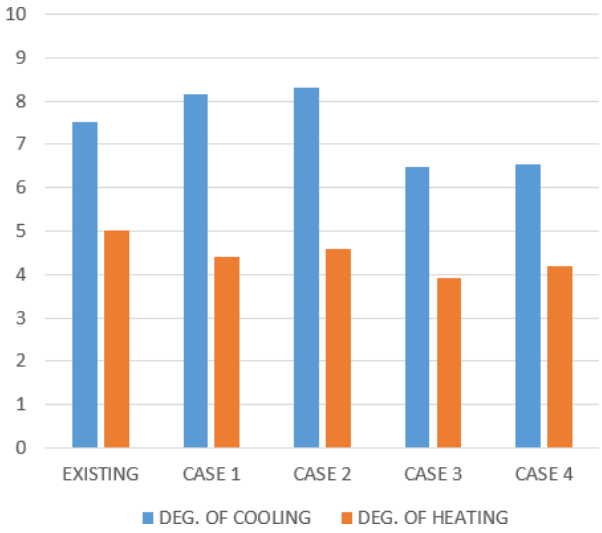


Figure 18: Degrees of cooling and heating – case 1-4

Table 3: Results of Case 5 to Case 8 (geometry modifications with case 2 condition)

CONDITIONS	HOT OUTLET [°C]	COLD OUTLET [°C]
CASE 5	46.838	29.600
CASE 6	46.414	29.675
CASE 7	45.913	29.869
CASE 8	45.745	29.885

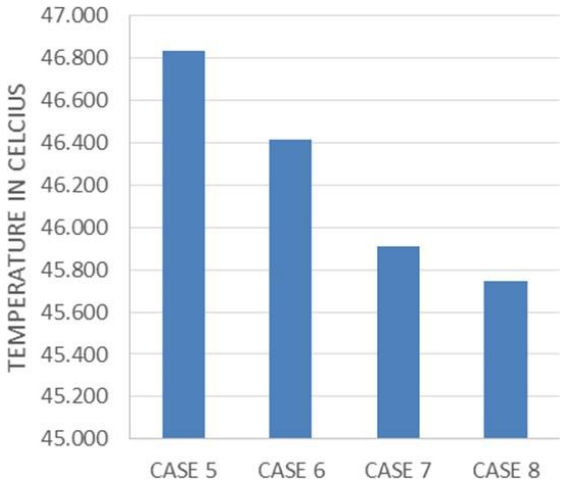


Figure 19: Hot water outlet temperatures – case 5-8

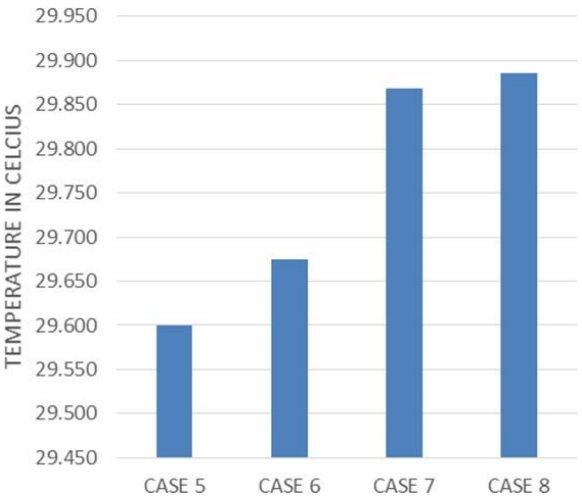


Figure 20: Cold water outlet temperatures case 5-8

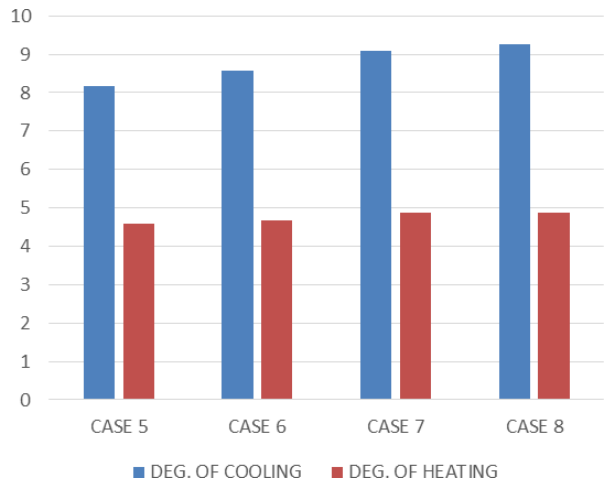


Figure 21: Degrees of cooling and heating – case 5-8

From the actual experimental data, the third serial data was taken for analysis and the obtained results are 47.5 and 30.0o C, whereas for the same data, the CFD results obtained are, 46.845 and 29.393o C (case 1). This proves that, the CFD results are comparatively matching with the experimental results. Hence on further analysis, it was found that the case 2 operating condition was having better results than the other three cases.

Using the case 2 operating conditions, the geometry of the heat exchanger was modified and analysed. On observation it was found that, the geometry with case 2 and slots is having slight better performance. On further improvement activity, the baffle gap was reduced and analysed. This shows that the performance increases as the volume of water inside the shell increases, but not beyond a limit as it affects the discharge issues. So the gap of 15mm was maintained and for this condition, the baffles are analysed for spherical dimples instead of slots. This condition results proves that the output is much better than the other operating conditions. The spherical dimples are used as an innovative idea.

Based on the above results, it was found that the modification of baffles with spherical dimples are having better performance than all other cases. In addition to this, an attempt was made to check the performance by changing the tube's cross section to rectangular from circular. The same was done and its performance results are plotted below.

Table 4: Results of Case 8 and Case 9

CONDITIONS	HOT OUTLET [°C]	COLD OUTLET [°C]
CASE 8	45.745	29.885
CASE 9	45.363	30.377

From the above it can be said that, the change in cross section of the tube is having considerable amount of impact in the performance of the system. On comparing the results, the proposed rectangular shape along with the spherical dimples is having better performance than the other cases.

CONCLUSION

The requirement of the industry was to analyse their heat exchanger for the using CFD tools and compare the results and if the results are satisfied, then the development of the system have to be carried out. This was taken as the problem statement and the actual setup was modelled and analysed under the actual boundary conditions to validate the CFD results. The CFD outputs have justified that the actual results and software results are mere co-incidence based on some assumptions.

Further the operating conditions of the system was varied and analysed to find the best effective condition. The case 2 condition was found to be the best ant it was suggested. Now based on this case 2 results, the modifications in baffle design were conducted and it was analysed under same boundary conditions. The baffle designs with spherical dimples was found to be having better performance compared to the other design suggestions.

Also, an initiative was made to check the performance by changing the cross section of the tubes for the design having better performance. This was implemented using the rectangular cross sections for the tube and the analysis was conducted and the results were found to be better than the all other proposed cases.

Hence, it can be concluded that, the rectangular tube is having better performance than the circular tube and the modifications in the baffle designs can be implemented to get the better performance.

REFERENCES

[1] Vindhya Vasiny Prasad Dubey, Raj Rajat Verma, Piyush Shanker And A.K.Srivastava, "Performance Analysis Of Shell & Tube Type Heat Exchanger Under The Effect Of Varied Operating Conditions", Journal Of Mechanical And Civil Engineering (Iosr-Jmce), Volume 11, Issue 3 Ver. Vi (May- Jun. 2014), Pp 08-17.

[2] Vikas Kumar, D. Gangacharyulu, Parlapalli Ms Rao And R. S. Barve, "Cfd Analysis Of Cross Flow Air To Air Tube Type Heat Exchanger".

[3] 3. Avinash D Jadhav, Tushar A Koli, "Cfd Analysis Of Shell And Tube Heat Exchanger To Study The Effect Of Baffle Cut On The Pressure Drop", International Journal Of Research In Aeronautical And Mechanical Engineering, 2321-3051i.

[4] 4. Hetal Kotwal, D.S Patel, "Cfd Analysis Of Shell And Tube Heat Exchanger- A Review", International

Journal Of Engineering Science And Innovative Technology, Volume 2, Issue 2, March 2013

[5] Alok Vyas, Mr. Prashant Sharma, "An Experimental Analysis Study To Improve Performance Of Tubular Heat Exchangers", Alok Vyas Et Al Int. Journal Of Engineering Research And Applications, Vol. 3, Issue 6, Nov-Dec 2013, Pp.1804-1809.

[6] Apu Roy, D.H.Das, "Cfd Analysis Of A Shell And Finned Tube Heat Exchanger For Waste Heat Recovery Applications", International Journal Of Mechanical & Industrial Engineering, Volume-1 Issue-1, 2011

Date: 18 | Dec-2019

"I.C. Engine Cylinder Fins Transient Thermal Analysis By Using ANSYS Software"

Satish Lingoji1, Nagaraju Dappu2

1Research Scholar, Mechanical Engineering Department MRCE Hyderabad (TG.)
2Research Scholar, Mechanical Engineering Department MRCE Hyderabad (TG.)

ABSTRACT - *The cooling mechanism of the air cooled engine is mostly dependent on the fin design of the cylinder head and block. Cooling fins are used to increase the heat transfer rate of specified surface. Engine life and effectiveness can be improved with effective cooling. The main aim of the project is to study and comparing with 100 cc Hero Honda Motorcycle fins and analyzes the thermal properties by varying geometry, material and thickness. Parametric models of cylinder with fins have been developed to predict the transient thermal behavior. Presently Material used for manufacturing the models is aluminum alloy 6063 which has thermal conductivity of 200W/mk. We are analyzing the designed models by taking the thermal temperature of 100 degrees centigrades. The energy transfers from the combustion chamber of an internal combustion engines are dissipate in three different ways. Transient thermal analyses were performed for actual and proposed design of engine cylinder in order to optimize geometrical parameters and enhanced heat transfer from the IC engine. Result revel that the proposed design of IC engine has better performance and heat transfer rate from the heating zone in the IC engine that is why the result of present work is more concentrate on it and also proposed replacement of new design by Using ANSYS 17.0 software.*

Key words: *Internal Combustion engine, transient thermal analysis, CATIA, ANSYS, Heat Transfer, Aluminum*

I. INTRODUCTION

Most combustion engines or internal combustion engines are fluid cooled using either air (aeriform fluid) or a liquid agent like water run through a device (radiator) cooled by air. In air cooling system, heat is carried out or driven away by the air flowing over and round the cylinder. Here fins are sew the plate and cylinder barrel which offer extra heat conductive and heat radiating surface. In water cooling system of cooling engines, the cylinder walls and heads are given jacket Cooling fins facilitate keep Chevrolet potential unit battery at ideal temperature we all know that just in case of internal combustion (IC) engines, combustion of air and fuel takes place within the engine.

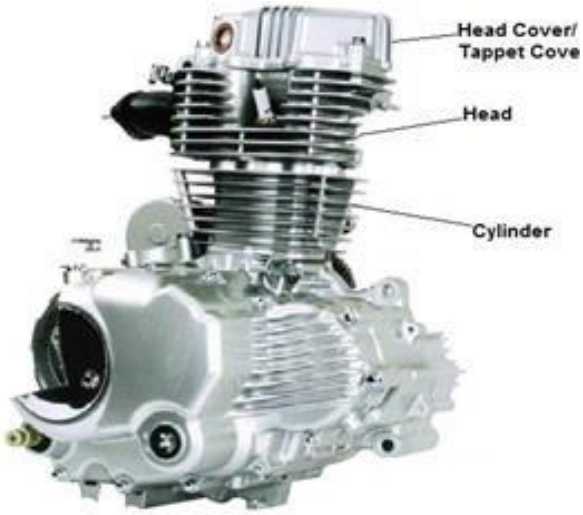


FIG. 1

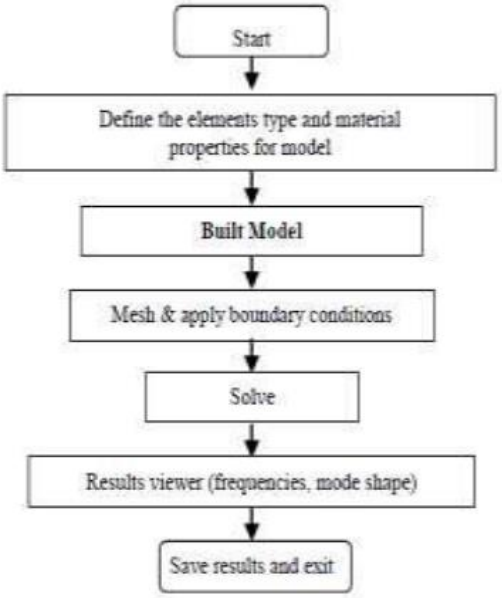


FIG. 2

Types of Cooling System

There are primarily Two types of cooling systems:

1. Air cooled system, and
2. Water cooled system.

II. METHODOLOGY

Step 1: Aggregation data and information associated with cooling fins of IC engines.
Step 2: A completely parametric model of the cylinder block with fin is formed in CATIA ver 5.0 software package.
Step 3: Model obtained in Step a pair of is analyzed using ANSYS17.0 (Workbench), to get the warmth or heat rate , thermal gradient and nodal temperatures.
Step 4: Manual calculations are done.
Step 5: Finally, we tend to compare the results obtained from ANSYS and manual calculations for completely different material, shapes and thickness.

FIN EQUATION

The purpose of fins in IC engine is to boost convective heat transfer from engine. The first purpose behind the operation of fins is to boost the effective heat transfer space from the surface. A balance of energy is performed on this component within which it's assumed that the component is at constant and uniform temperature of T.

Where C_1 and C_2 are constant that can be determined from the boundary conditions.

Case I: The fin is very long and the temperature at the end of the fin is approach to the temperature of the surrounding T_∞.

Conductive heat transfer at the base of fin, according to Fourier's law

$$Q_{Fin} = -kA\left(\frac{dT}{dx}\right)_{x=0}$$

$$Q_{Fin} = \sqrt{hPkA}(T_0 - T_\infty)$$

K = 200 w/mk = 0.2 w/mm k

h = 5 x $10^{-0.006}$ w/mm² k

T_0 = 1000 ⁰C

T_∞ = 25 ⁰C

$$Q_{Fin} = \sqrt{hPkA}(T_0 - T_\infty)$$

Width of fins = 1.5 mm

Thickness of fins = 1 mm

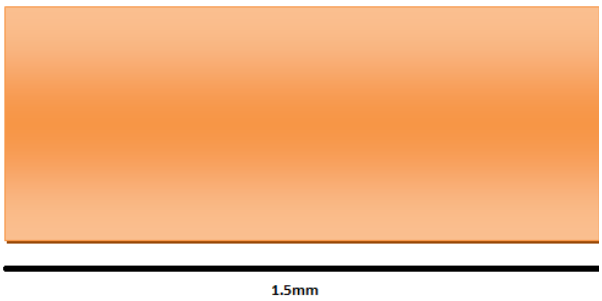


FIG. 3

Q = √ 5x10-0.006 x 2 (1.5 + 1) x 0.2 x 1.5 x 1 x 975

Q = 2670 watt

Material properties and Boundary Conditions

Engine: 100 cc Hero Honda

Aluminum Alloy 6063

Thermal conduction K = 200 W/m-K=0.2 W/mm-K

Specific heat Cp = 0.9 J/g0C = 900 J/Kg-K

Density = a pair of 2.7 g/cc = 2700 kg/m^3= 0.0000027 kg/mm^3

Boundary Condition:

Ambient Temperature: 250

Cylinder Internal temp. = 1000 0C

Heat Flux = 22 W/mm2

Film constant worth = 5x10-006 w/mm2 0C

III. SIMULATION

EXITING MODEL

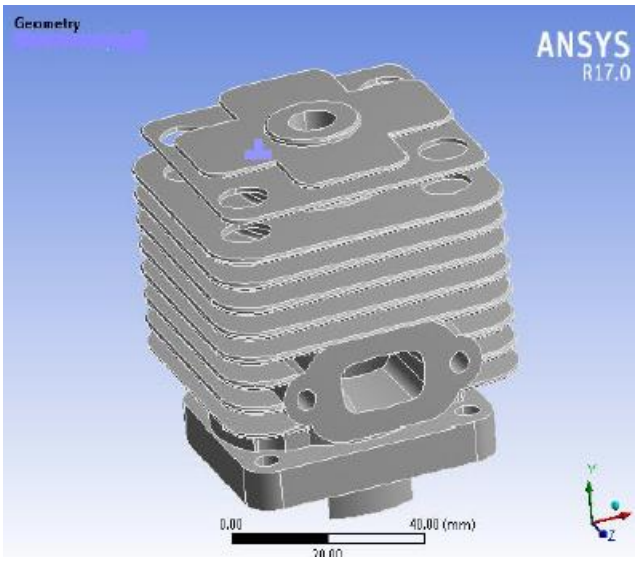


Fig.04

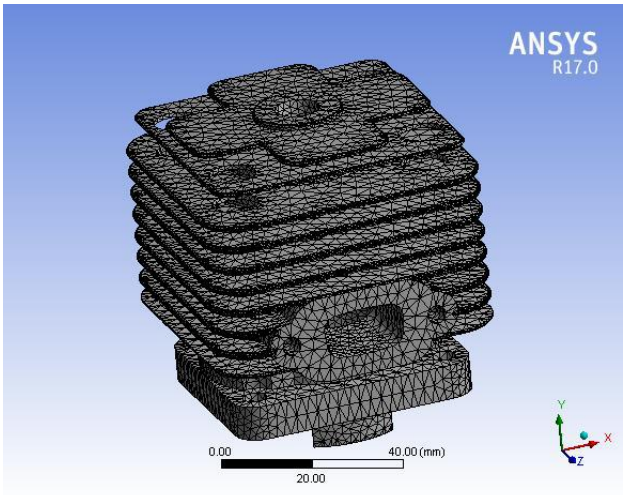


Fig.05

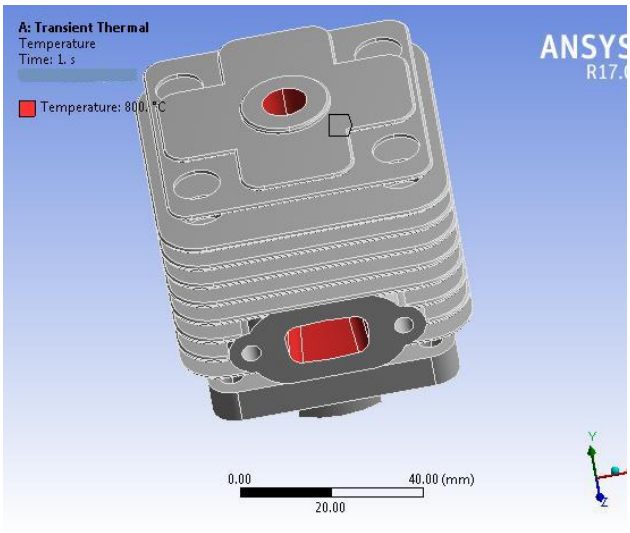


Fig.06

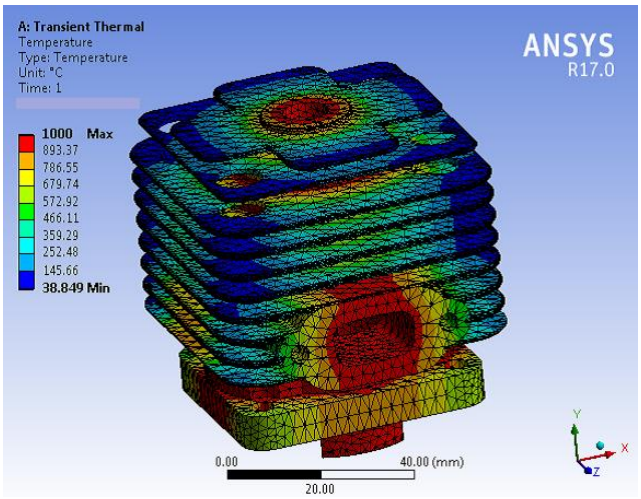


Fig.08

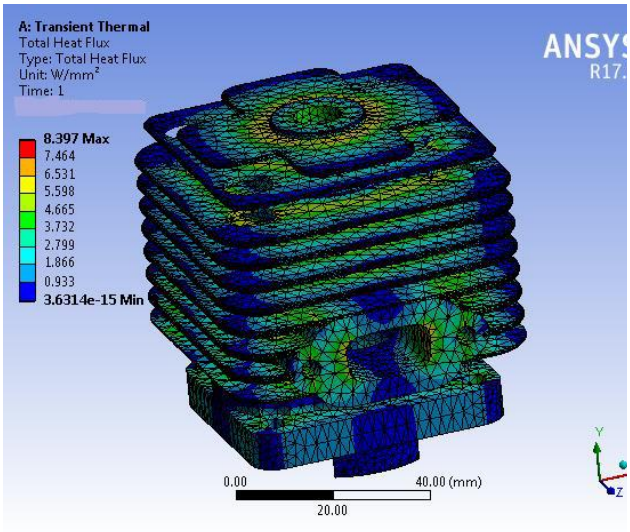


Fig.07

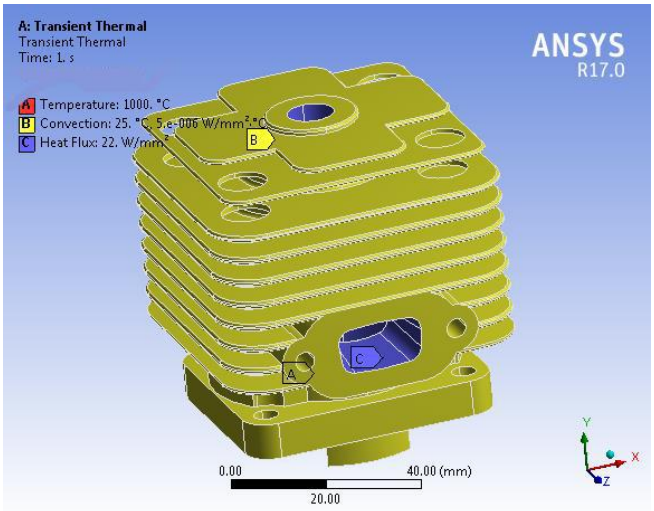


Fig.09

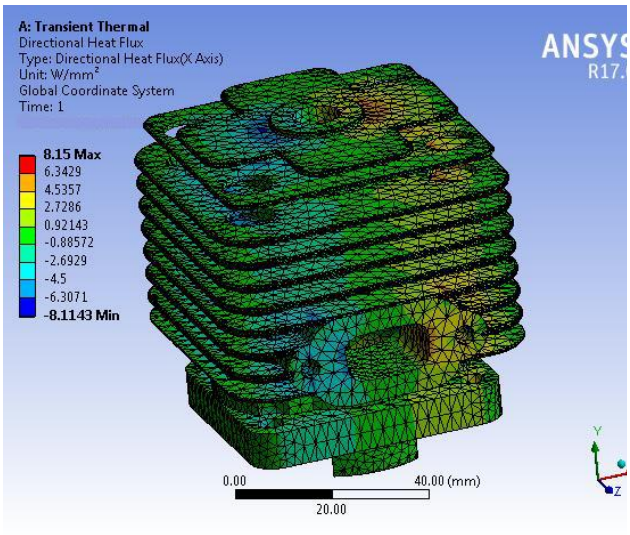


Fig.10

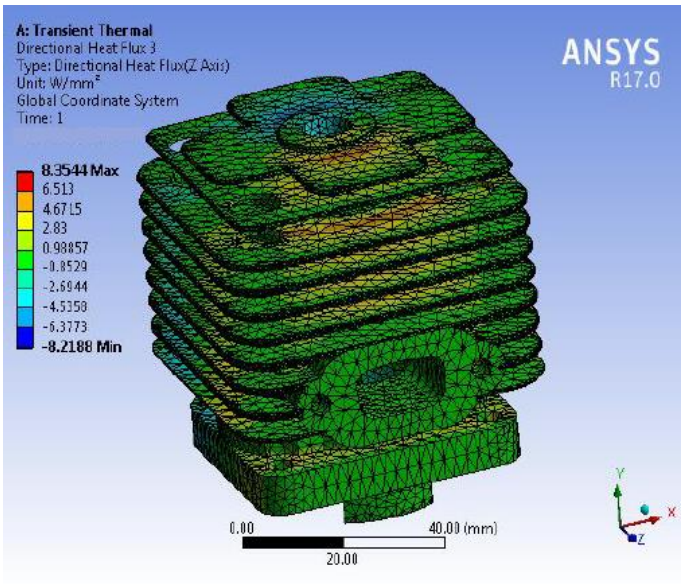


Fig.12

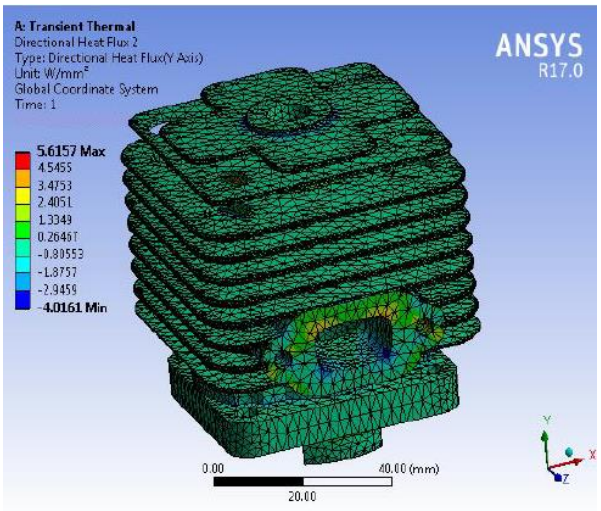


Fig.11

PROPOSED MODEL

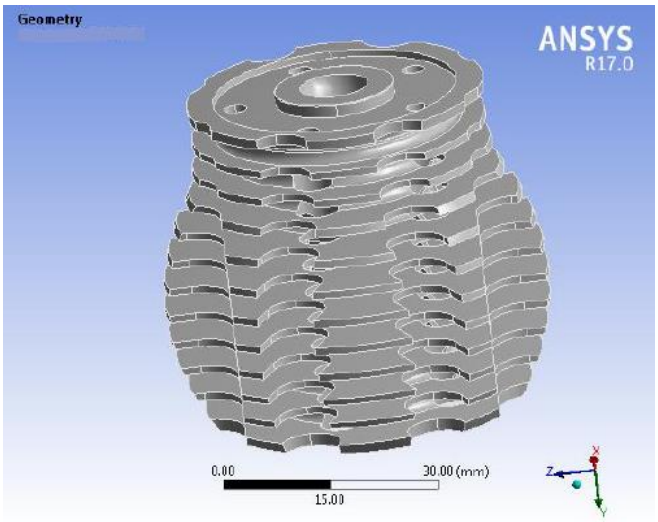


Fig.13

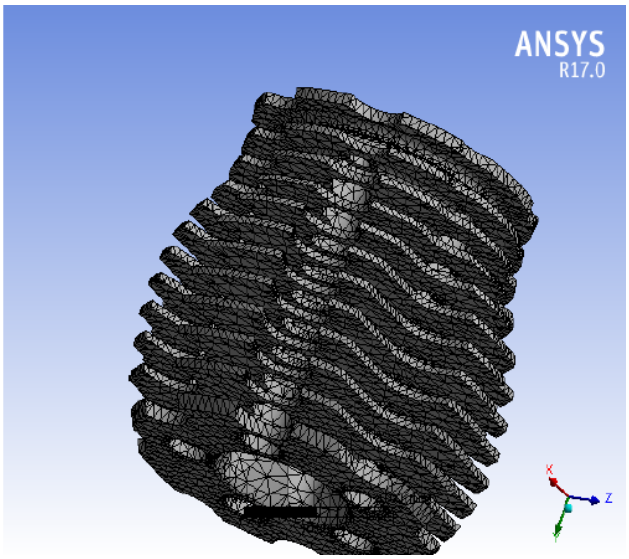


Fig.14

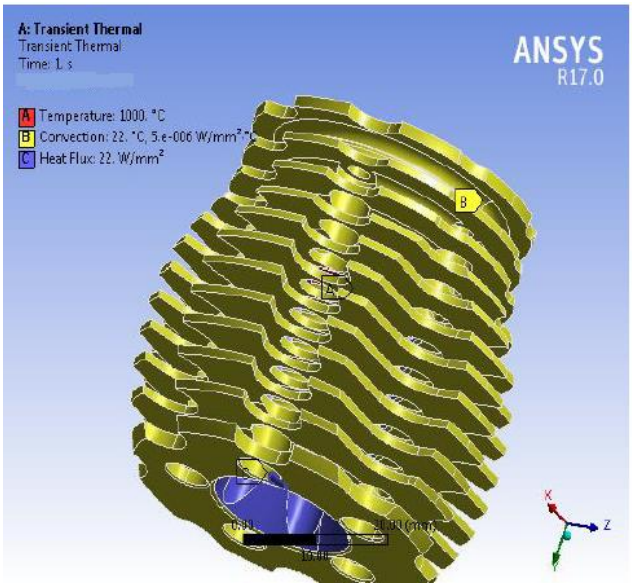


Fig.16

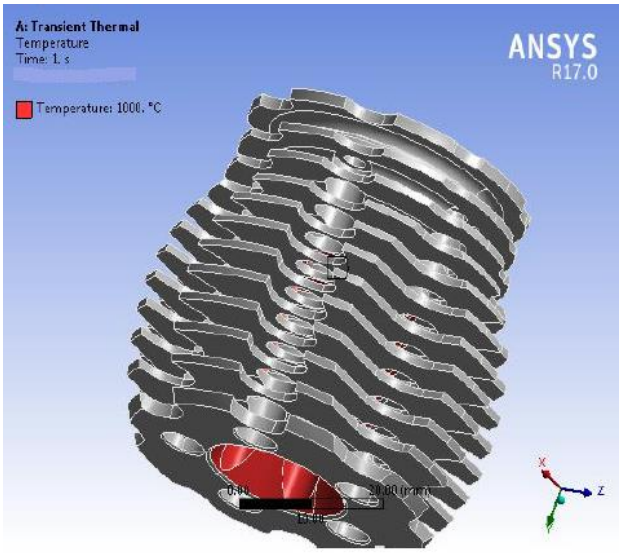


Fig.15

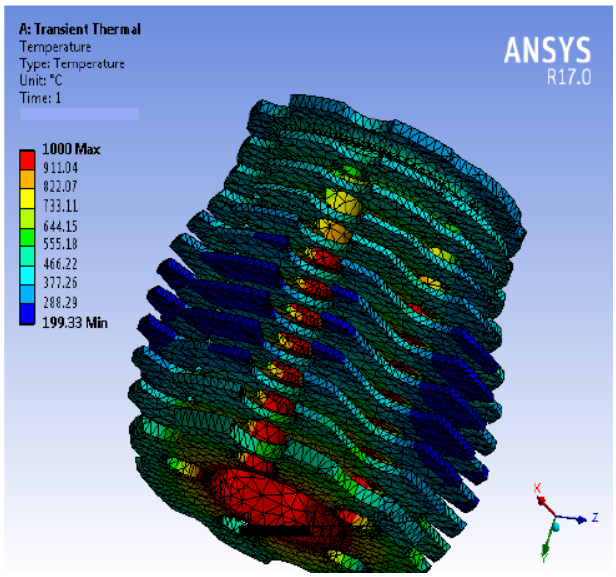


Fig.17

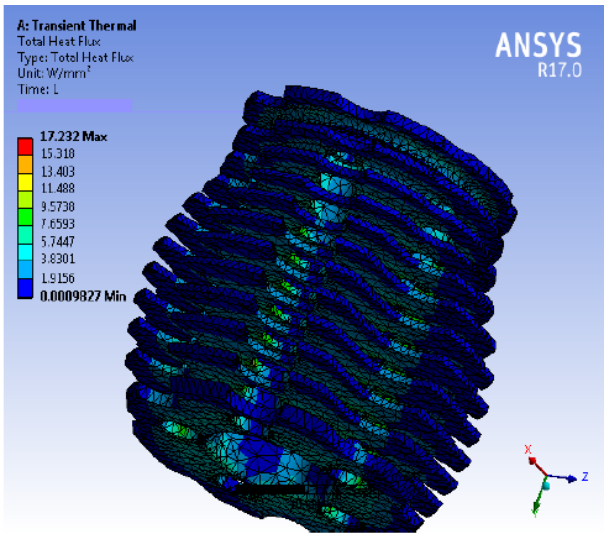


Fig.18

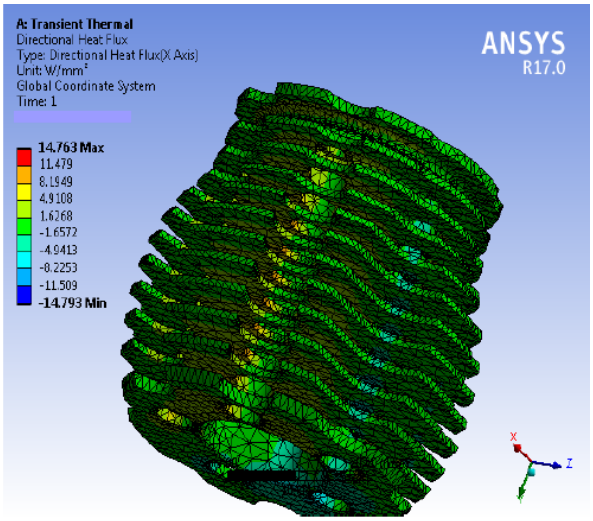


Fig.19

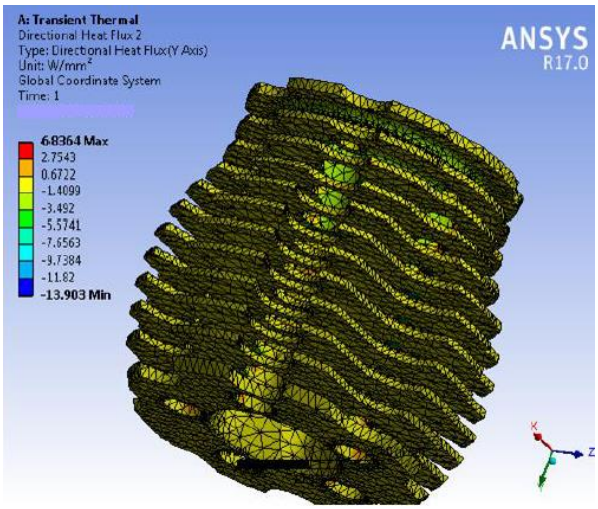


Fig.20

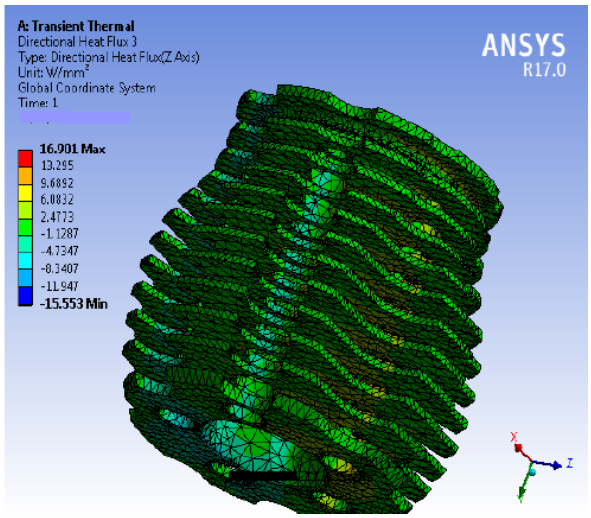


Fig.21

Table.1

S.NO	Temp(^{0}C)	Exiting Model	Proposed Model
Temp(^{0}C)	*Max*	*1000*	*1000*
	Min	*38.84*	*199.33*
Total Heat (w/mm2)Flux	*Max*	*8.397*	*17.2*
	Min	*3.63E-15*	*9.83E-04*
Directional Heat Flux(w/mm2) (X Direction)	*Max*	*8.15*	*14.76*
	Min	*-8.11*	*-14.79*
Directional Heat Flux(w/mm2) (Y Direction)	*Max*	*5.6*	*6.8*
	Min	*-4.0*	*-13.9*
Directional Heat Flux(w/mm2) (Z Direction)	*Max*	*8.3*	*16.9*
	Min	*-8.2*	*-15.5*

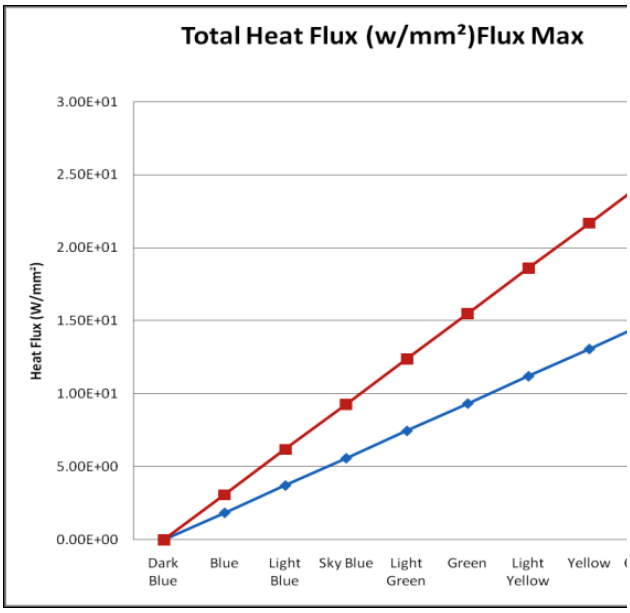


Fig.22

IV. RESULT & DISCUSSION

The transient thermal analysis ware performed using an analytical software package ANSYS worktable supported finite volume analysis. The consequences of various vital geometrical parameters for the transient natural convective heat transfer rate from each actual and projected style of engine.
Transient thermal analyses were performed for actual and projected design of engine cylinder so as to optimize geometrical parameters and increased heat transfer from the IC engine. Within the present work transient thermal analysis is performed on actual style and additionally on two completely different geometrical styles at close temperature 25 ^{0}C. The subsequent points are recognized within the variety of conclusive statements that are as follows.
1. The results of transient thermal analysis of actual design of engine cylinder at close temperature 26oC indicates the utmost temperature is 1000 oC and minimum temperature is 199.3 oC, most enthalpy flux generated is 17.2 W/mm2 and minimum heat flux generated is 9.83E-04 W/mm2, most Directional heat flux in Y-direction generated is 13.2 W/mm2 and minimum Directional heat flux generated is -7.46 W/mm2
2. The results of transient thermal analysis of proposed model of engine cylinder at close temperature 25 ^{0}C indicates the utmost temperature is 1000 ^{0}C and minimum temperature is 167.8 ^{0}C, most total heat flux generated is 27.8 W/mm2 and minimum heat flux generated is 0.002828 W/mm2, the utmost directional heat flux in X-direction generated is 14.76 W/mm2 and minimum

directional heat flux generated is -14.79 /mm2 , directional heat flux in Y-direction generated is 6.8 W/mm2 and minimum Directional heat flux generated is -13.9W/mm2 and directional heat flux in Z-direction generated is 16.9 W/mm2 and minimum Directional heat flux generated is -15.5 W/mm2.

V. CONCLUSION

During this paper we've got designed a cylinder fin body used n a 100cc Hero Honda motorbike and 3D modeling software package CATIA ver 5.0 and used material for fin body is metallic element alloy fins and internal core with gray forged iron. We have a tendency to are commutation with metallic element alloy 6063 for entire body. the form of the fin is rectangular; we've got modified the form with circular geometry formed. To summarize this conclusion, the projected style of IC engine has higher performance and heat transfer rate from the heating zone within the IC engine that's why the results of present work is additional focus on it and additionally projected replacement of recent style.

VI. FUTURE SCOPE:

The aim of gift work to extend heat transfer rate from the heating zone in IC engine, for that transient thermal analysis are performed on actual style of Hero Honda 110 CC single cylinder engine. There are some doable future works which can be doable for more analysis;
1. Radiation analysis also can be performed for constant work.
2. Altogether sorts of analysis within the present work the fabric used for casting is Al alloy; another material may used.
3. CFD analysis also can be done to grasp air flow round the casting.

REFERENCES

[1]**Shubham Shrivastava & Shikar Upadhyay** "Thermal Analysis of IC Engine Cylinder Block with Fins Perpendicular to the Axis of Piston Movement" International Journal of Mechanical and Industrial Technology ISSN 2348-7593 (Online) Vol. 3, Issue 2, pp: (139-149), Month: October 2015 - March 2016.

[2] Vinay Kumar Attar & Himanshu Arora "Transient Thermal Analysis of Internal Combustion Engine Piston in Ansys Workbench By Finite Element Method" International Journal of Engineering Sciences & Research Technology, ISSN: 2277- 9655 Attar* et al., 5(6): June, 2016.

[3] Chidiebere Okeke-Richard & Sunny Sharma " Thermal Analysis and Comparison of Cylinder Blocks of 4S, SITwo Wheeler Engine Using ANSYS" International Journal of Innovative Science, Engineering & Technology, ISSN 2348 – 7968 Vol. 3 Issue 5, May 2016.

[4] G.Bahadur Vali & Krishna Veni "Design/Modeling and Thermal Analysis on Cylinder Head of I.C Engine" International Journal & Magazine of Engineering, Technology, Management and Research, ISSN No. 2348- 4845 Volume No.03 (2016), Issue No.05 (may).

[5] Abhishek Mote at el "Analysis of Heat transfer through fins of an IC Engine using CFD" International Research Journal of Engineering and Technology (IRJET) e- ISSN: 2395 -0056, p- ISSN: 2395-0072 , Volume: 03 Issue: 04 | Apr-2016.

[6] KM Sajesh, Neelesh Soni and Siddhartha Kosti "Design Modification and Heat Transfer Analysis of Air Cooled Rectangular Fin Engine" International Journal of Recent Scientific Research ISSN: 0976-3031 Volume: 7(3) March - 2016.

[7] Mr. Manir Alam & Mrs. M. Durga Sushmitha "Design and Analysis of Engine Cylinder Fins of Varying Geometry and Materials" International Journal of Computer Engineering In Research Trends, ISSN (O): 2349-7084 Volume 3, Issue 2, February- 2016, pp. 76-80.

[8] Swati Saini & Kagdi Dhruvin Nileshkumar "Comparison of CFD Analysis of Heat Transfer Augmentation & Surface Heat Transfer Co efficient of I. C. Engine Fins at Different wind Velocities with Experimental Investigation" International Journal for Scientific Research & Development| Vol. 4, Issue 03, 2016 | ISSN (online): 2321- 0613.

[9] P.T. Nitnaware & Prachi S. Giri "Design Optimization Of An Air Cooled Internal Combustion Engine Fin Using CFD" Journal of Multidisciplinary Engineering Science and Technology (JMEST) ISSN: 3159- 0040 Vol. 2 Issue 11, November - 2015.

[10] H.Sumithra & B. Sandhya Rani "Design Optimization and Transient Thermal Analysis of Four Stroke Petrol Engine Cylinder Head" International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE) ISSN: 0976- 1353 Volume 18 Issue 2 – NOVEMBER 2015.

[11] M Syamala Devi1, E Venkateswara Rao & K Sunil Ratna Kumar"Optimization of Engine Cylinder Fin by Varying its Geometry and Material" International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064, Paper ID: SUB159073 , Volume 4 Issue 10, October 2015.

[12] Athirah Abdul Aziz, Adlansyah Abd Rahman and Abdul Aziz Hassan "Design of Customized Modular Cylinder Head for Si-Engine" ARPN Journal of Engineering and Applied Sciences ISSN 1819-6608, VOL. 10, NO. 17, SEPTEMBER 2015.

[13] Gokul Karthik "Design Modification and Analysis of Fins in Two Wheeler Engines by CFD" International Journal of Research in Electrical and Electronics Technology (IJREET) Volume 2 - Issue 1 August 2015.

[14] Shamim Alam & Dr. Gaurav Tiwari "Increase the Efficiency of Internal Combustion Engine Fin by Replacing the Conventional Fin with Parabolic Fin" International Journal of Analytical, Experimental and Finite Element Analysis (IJAEFEA), Issue. 2, Vol. 2, June 2015. e-ISSN: 2394-5141, ISSN: 2394-5133, pp 80-84.

[15] Ashwin Shridhar, Asokan Ram Deepak, S Jayaraj "Int. Journal of Engineering Research and Applications " Conjugate Heat transfer Analysis of helical fins with airfoil cross section and its comparison with existing circular fin design for air cooled engines employing constant rectangular cross-section, ISSN : 2248- 9622, Vol. 5, Issue 6, (Part -5) June 2015, pp.88

Optimization of MIG Welding Parameters for Improving Strength of Welded Joints

SRIKANTH MUTHAMALA
M.TECH (THERMAL ENGINEERING)
MECHANICAL ENGINEERING DEPARTMENT
MALLA REDDY COLLEGE OF ENGINEERING

Abstract: **The problem that has faced the manufacturer is the control of the process input parameters to obtain a good welded joint with the required weld quality. Traditionally, it has been necessary to study the weld input parameters for welded product to obtain a welded joint with the required quality. To do so, requires a time-consuming trial and error development method. Then welds are examined whether they meet the requirement or not. Finally, the weld parameters can be chosen to produce a welded joint that closely meets the joint qualities. Also, what is not achieved or often considered is an optimized welding parameters combination, since welds can often be formed with very different parameters. In other words, there is often a more ideal welding input parameters combination, which can be used. In this thesis, the influence of welding parameters like welding current, welding voltage, welding speed on ultimate tensile strength (UTS) of AISI 1050 mild steel material during welding. A plan of experiments based on Taguchi technique has been used. An Orthogonal array, signal to noise (S/N) ratio and analysis of variance (ANOVA) are employed to study the welding characteristics of material & optimize the welding parameters. The result computed is in form of contribution from each parameter, through which optimal parameters are identified for maximum tensile strength. From this study, it is observed that welding current and welding speed are major parameters which influence on the tensile strength of welded joint.**

Key words: **MIG Welding; Speed; Taguchi Method; Mini Tab; Tensile Test;**

I. INTRODUCTION

Welding is, at its core, simply a way of bonding two pieces of metal. While there are other ways to join metal (riveting, brazing and soldering, for instance), welding has become the method of choice for its strength, efficiency and versatility.

There are tons of different welding methods, and more are being invented all the time. Some methods use heat to essentially melt two pieces of metal together, often adding a "filler metal" into the joint to act as a binding agent. Other methods rely on pressure to bind metal together, and still others use a combination of both heat **and** pressure. Unlike soldering and brazing, where the metal pieces being joined remain unaltered, the process of welding always changes the work pieces.

WELDING TOOLS OF THE TRADE

The most basic welding rigs, for occasional use in a home workshop, can be had for under $100. Typically, these rigs are set up for **shielded metal arc welding**(**SMAW**), or **stick welding**. Many units only have an on/off switch in the way of controls, making them simple to operate. **Torch welding rigs** are small and easy to work with, which is part of why they're commonly used. These torches use oxyacetylene for the flame, along with a filler rod. But some rigs (like those used in laser-beam welding) are so expensive and complicated that they are only used in industrial applications.As for materials, some are much easier to weldthan others. Steel can be a great choice because of its strength, affordability and weld ability. As a rule, the stronger the steel, the harder it is to weld. Accordingly, several steel alloys were developed with welding in mind. Of course, almost any metal can be welded, including cast iron, bronze, aluminium and even titanium, although the latter requires a highly protected atmosphere because the metal is so reactive.

Whatever you're welding, remember: safety first. If you've ever seen welding in person, you can testify to the blinding brightness the process creates. Looking directly at a weld site without protection can produce what's known as **arc eye**, a painful inflammation of the cornea that feels like getting sand in your eye. No wonder that a good welder's mask is a prerequisite for any welding outfit.

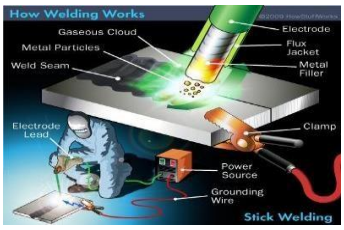


Fig1. welding process

THE PROCESS OF WELDING

Most welding done today falls into one of two categories: arc welding and torch welding.

Arc welding uses an electrical arc to melt the

work materials as well as filler material (sometimes called the welding rod) for welding joints. Arc welding involves attaching a grounding wire to the welding material or other metal surface. Another wire known as an electrode lead is placed on the material to be welded. Once that lead is pulled away from the material, an electric arc is generated. It's a little like the sparks you see when pulling jumper cables off a car battery. The arc then melts the work pieces along with the filler material that helps to join the pieces.

Feeding the filler into the welding joint takes steady hands and an eye for detail. As the rod melts, the welder must continuously feed the filler into the joint using small, steady, back-and-forth motions. These motions are what gives welds their distinctive appearance. Going too fast or slow, or holding the arc too close or far away from the material can create poor welds.

Shielded metal arc welding (SMAW or stick welding), gas metal arc welding (more commonly known as metal inert gas, or MIG, welding) and gas tungsten arc welding (frequently called tungsten inert gas, or TIG, welding) all exemplify arc welding.

These three common methods each offer unique advantages and drawbacks. Stick welding, for instance, is inexpensive and easy to learn. It's also slower and less versatile than some other methods. Oppositely, TIG welding is difficult to learn and requires an elaborate welding rig. TIG welding produces high-quality welds, however, and can weld materials that other methods can't.

Torch welding represents another popular welding method. This process typically uses an oxyacetylene torch to melt the working material and welding rod. The welder controls the torch and rod simultaneously, giving him or her a lot of control over the weld. While torch welding has become less common industrially, it's still frequently used for maintenance and repair work, as well as in sculptures (more on that later).

II. LITERATURE REVIEW

Gas Metal Arc Welding (GMAW), sometimes referred to by its subtypes Metal Inert Gas (MIG) welding or Metal Active Gas (MAG) welding, is a semi-automatic or automatic 0020 Arc welding process in which a continuous and consumable wire electrode and a shielding gas are fed through a welding gun. A constant voltage, direct current power source is most commonly used with GMAW, but constant current systems, as well as alternating current, can be used. In this research work an attempt was made to develop a response surface model to predict tensile strength of inert gas metal arc welded AISI 1040 medium carbon steel joints. The process parameters such as welding voltage, current, wire speed and gas flow rate were studied. The experiments were conducted based on a four-factor, three-level, face centred composite design matrix. The empirical relationship can be used to predict the yield strength of inert gas metal arc welded AISI 1040 medium carbon steel. Response Surface Methodology (RSM) was applied to optimizing the MIG welding process parameters to attain the maximum yield strength of the joint.

III. PROBLEM DESCRIPTION

Objective of the work

In this thesis, materials AISI 1050 Mild Steel are welded by varying process parameters welding speed, welding current and welding voltage. Effect of process current on the tensile strength of weld joint will be analysed.

EXPERIMENTAL PROCEDURE

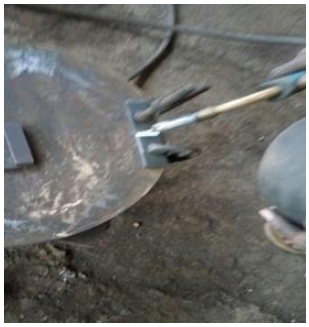
In this thesis, experiments are made to understand the effect of MIG welding parameters welding speed, welding current and welding voltage on output parameters such as hardness of welding, tensile strength of welding.

MIG welding experimental images

MIG welding machine

Work pieces (AISI 1050 STEEL)

Dumbbell shape work pieces for tensile test



Work pieces' setup

Welding process

For the experiment, welding parameters selected are shown in table.

The welding current and electrodes considered are

PROCESS PARAMETERS	LEVEL1	LEVEL 2	LEVEL 3
WELDING CURRENT (AMP)	180	230	280
WELDING SPEED (m. m/s)	200	300	400
WELDING VOLTAGE (V)	22	24	26

WELDING CURRENT (AMP)	WELDING SPEED (m. m/s)	WELDING VOLTAGE (V)
180	200	22
180	300	24
180	400	26
230	200	22
230	300	24
230	400	26
280	200	22
280	300	24
280	400	26

WELDING CURRENT (AMP)	WELDING SPEED (m. m/s)	WELDING VOLTAGE (V)	ULTIMATE TENSILE STRENGTH (MPa)
180	200	22	375
180	300	24	410
180	400	26	451.197
230	200	22	403
230	300	24	440.581
230	400	26	372
280	200	22	375.287
280	300	24	369
280	400	26	378

IV. TENSILE TEST REPORTS

INTRODUCTION TO TAGUCHI TECHNIQUE

- Taguchi defines Quality Level of a product as the Total Loss incurred by society due to failure of a product to perform as desired when it deviates from the delivered target performance levels.
- This includes costs associated with poor performance, operating costs (which changes as a product ages) and any added expenses due to harmful side effects of the product in use.

Taguchi Methods

- Help companies to perform the Quality Fix!
- Quality problems are due to Noises in the product or process system
- Noise is any undesirable effect that increases variability

- Conduct extensive Problem Analyses
- Employ Inter-Disciplinary Teams
- Perform Designed Experimental Analyses
- Evaluate Experiments using ANOVA and Signal-to noise techniques

TAGUCHI PARAMETER DESIGN FOR TURNING PROCESS

In order to identify the process parameters affecting the selected machine quality characteristics of turning, the following process parameters are selected for the present work: cutting speed (A), feed rate (B) and depth of cut (C). the selection of parameters of interest and their ranges is based on literature review and some preliminary experiments conducted.

Selection of Orthogonal Array

The process parameters and their values are given in table. It was also decided to study the two – factor interaction effects of process parameters on the selected characteristics while turning. These interactions were considered between cutting speed and feed rate (AXB), feed rate and depth of cut (BXC), cutting speed and depth of cut (AXC).

PROCESS PARAMETERS	LE VEL1	LEVEL 2	LEVEL 3
WELDING CURRENT (AMP)	180	230	280
WELDING SPEED (m. m/s)	200	300	400
WELDING VOLTAGE (V)	22	24	26

Using randomization technique, specimen was turned and cutting forces were measured with the three – dimensional dynamometer. The experimental data for the cutting forces have been reported in Tables. Feed and radial forces being 'lower the better' type of machining quality characteristics, the S/N ratio for this type of response was and is given below:

$$\text{S/N ratio} = -10 \log\left[\frac{1}{n}(y_1^2 + y_2^2 + ... + y_n^2)\right] \quad \ldots (1)$$

Where $y_1, y_2, \ldots, y_n$ are the responses of the machining characteristics for each parameter at different levels.

TAGUCHI ORTHOGONAL ARRAY

WELDING CURRENT (AMP)	WELDING SPEED (m. m/s)	WELDING VOLTAGE (V)
180	200	22
180	300	24
180	400	26
230	200	22
230	300	24
230	400	26
280	200	22
280	300	24
280	400	26

OBSERVATION

The following are the observations made by running the experiments. The ultimate tensile strength observed.

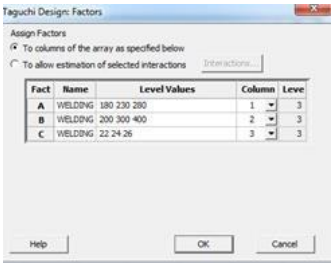
UTS (MPa)
375
410
451.917
403
440.581
372
375.287
369
378

OPTIMIZATION OF ULTIMATE TENSILE STRENGTH USING MINITAB SOFTWARE

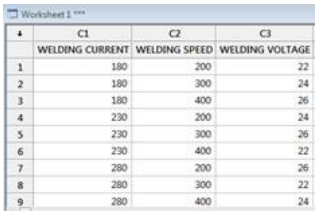
Design of Orthogonal Array

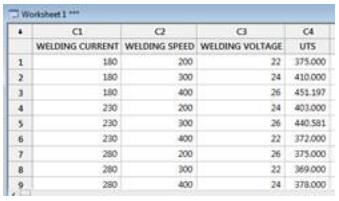
First Taguchi Orthogonal Array is designed in Minitab17 to calculate S/N ratio and Means which steps is given below:

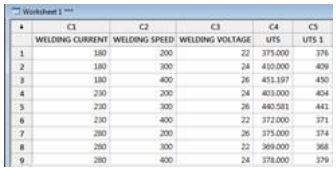
FACTORS



OPTIMIZATION OF PARAMETERS



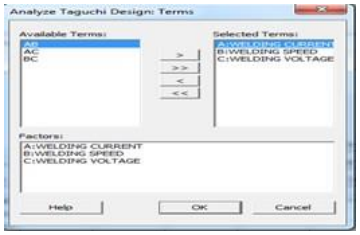


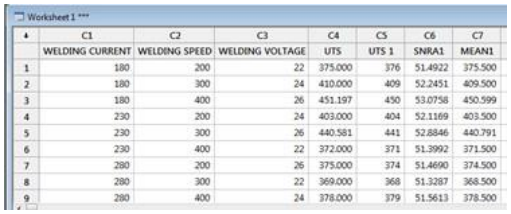


	C1	C2	C3	C4	C5
	WELDING CURRENT	WELDING SPEED	WELDING VOLTAGE	UTS	UTS 1
1	180	200	22	375.000	376
2	180	300	24	410.000	409
3	180	400	26	451.197	450
4	230	200	24	403.000	404
5	230	300	26	440.581	441
6	230	400	22	372.000	371
7	280	200	26	375.000	374
8	280	300	22	369.000	368
9	280	400	24	378.000	379

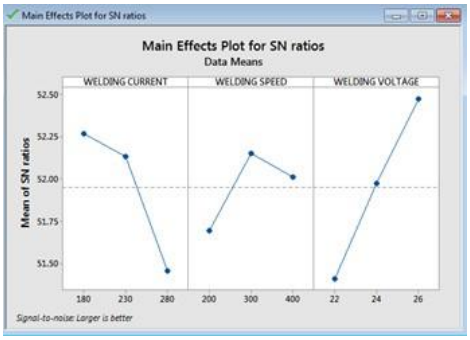
Analyse Taguchi Design – Select Responses

Terms





	C1	C2	C3	C4	C5	C6	C7
	WELDING CURRENT	WELDING SPEED	WELDING VOLTAGE	UTS	UTS 1	SNRA1	MEAN1
1	180	200	22	375.000	376	51.4922	375.500
2	180	300	24	410.000	409	52.2451	409.500
3	180	400	26	451.197	450	53.0758	450.599
4	230	200	24	403.000	404	52.1169	403.500
5	230	300	26	440.581	441	52.8846	440.791
6	230	400	22	372.000	371	51.3992	371.500
7	280	200	26	375.000	374	51.4690	374.500
8	280	300	22	369.000	368	51.3287	368.500
9	280	400	24	378.000	379	51.5613	378.500



S/N ratio plot

Means plot

V. RESULTS

Taguchi method stresses the importance of studying the response variation using the signal–to–noise (S/N) ratio, resulting in minimization of quality characteristic variation due to uncontrollable parameter. The cutting force is considered as the quality characteristic with the concept of "the larger-the-better". The S/N ratio for the larger-the-better is:

$S/N = -10 * \log(\Sigma(Y^2)/n))$

Where n is the number of measurements in a trial/row, in this case, n=1 and y is the measured value in a run/row. The S/N ratio values are calculated by taking into consideration above Eqn. with the help of software Minitab 17.

The force values measured from the experiments and their corresponding S/N ratio values are listed in Table

	C1	C2	C3	C4	C5	C6	C7
	WELDING CURRENT	WELDING SPEED	WELDING VOLTAGE	UTS	UTS 1	SNRA1	MEAN1
1	180	200	22	375.000	376	51.4922	375.500
2	180	300	24	410.000	409	52.2451	409.500
3	180	400	26	451.197	450	53.0758	450.599
4	230	200	24	403.000	404	52.1169	403.500
5	230	300	26	440.581	441	52.8846	440.791
6	230	400	22	372.000	371	51.3992	371.500
7	280	200	26	375.000	374	51.4690	374.500
8	280	300	22	369.000	368	51.3287	368.500
9	280	400	24	378.000	379	51.5613	378.500

VI. CONCLUSION

The experiment designed by Taguchi method fulfils the desired objective. Fuzzy interference system has been used to find out the ultimate tensile strength. The all possible values of have been calculated by using MINITAB 17.0 software. Analysis of variance (ANOVA) helps to find out the significance level of each parameter. The optimum value was predicted using MINITAB-17 software.

The welding parameters are Welding current, welding voltage and welding speed for MIG welding of work piece AISI1050 steel. In this work, the optimal parameters of welding speed are 200m.m/s, 300 m.m/s & 400 m.m/s, welding current are 180, 230 &280 amps, and welding voltage are 22, 24 &26 volts. Experimental work is conducted by considering the above parameters. Ultimate tensile strength validated experimentally.

The experimental results confirmed the validity of the used Taguchi method for enhancing the welding performance and optimizing the welding parameters in MIG welding at welding speed 400 m.m/s, welding voltage 26 volts and welding current 180 amps

VII. REFERENCES

[1]. Optimization of mig welding process parameters to predict maximum yield strength in aisi 1040 ajit hooda1*, ashwani dhingra2 and satpal sharma2

[2]. A Review on Optimization of MIG Welding

Parameters using Taguchi's DOE Method Satyaduttsinh P. Chavda1, Jayesh V.Desai2 , Tushar M.Patel3

[3]. Optimization of the process parameters for mig welding of aisi 304 and is 1079 using fuzzy logic method prasenjit mondal1, dipankar bose2

[4]. Optimization of MIG welding Process Parameter using Taguchi Techniques Kapil B. Pipavat1 , Dr. Divyang Pandya2 , Mr. Vivek Patel3

[5]. Optimization of MIG Welding Parameters for Hardness of Aluminium Alloys Using Taguchi Method Vineeta Kanwal1 , R S Jadoun2

[6]. Parameters Optimization for Tensile Strength & Hardness of MIG Welding Joint of HSS & Mild Steel by Using Taguchi Technique Er. Rahul Malik1 , Er. Surjeet Gahlot2 , Dr. S.K. Jarial3

[7]. A Review on Parametric Optimization of MIG Welding Parameters by using Various Optimization Techniques 1Pranesh B. Bamankar Asst. Professor Arvind Gavali College of Engg. Satara 2Amol Chavan Student Arvind Gavali College of Engg. Satara Tushar Phadtare Student Arvind Gavali College of Engg. Satara

Additive fabrication in Automotive sector

Abstract

Automotive sector faces new challenges every day, new design trends and technological deployments from research push companies to develop new models and facelifts in short term, requiring new tools or tool reshaping. The Automotive sector is one of the most competitive business areas where time-to-market decrease plays an important role. Additive fabrication is the solution which enables the flexible production of customized products without significant impact on costs and lead time. Automotive companies developing new models and facelifts every day, pushed by new design trends and technological evolution where aesthetics, aerodynamics, safety, and weight reduction of the vehicle are key issues. Therefore, new tools or tool reshaping for new components is required, for body panels or other technical components. Digitalization helps the Automotive sector to turn their ideas into successful vehicle faster and more efficiently. Automobile manufacturers can increase the efficiency of their research and development processes, enabling them to get their products on the market with less time and efficiently.. An important parameter of using metal AM processes in the automotive sector is fabricating complex lightweight structures which at the same time possess good rigidity. The weight of the automotive parts can be reduced significantly by using the ability of AM processes to maximum advantage to produce parts with complex geometries while maintaining relative strengths. This paper provides a review of today and future of additive fabrication in Automotive sector.

Keywords: Automotive sector, Additive fabrication, 3D printing, STL, CAD.

1. Introduction

Additive fabrication commonly known as 3D printing allows the direct conversion of design construction files into fully functional objects. It is a process of joining materials to make object from 3D model data usually layer upon layer. In this the material is joined or solidified under computer control to create a three-dimensional object, with material being added together (such as liquid molecules or powder grains being fused together) typically layer by layer.

Vehicle manufacturers have been at the forefront of implementing additive fabrication technology. One of the most prolific additive fabrication applications in the automotive world is in rapid prototyping. This is also among the oldest uses of the technology, with some large auto manufacturers having prototyped parts with 3D printers for more than 20 years. Beyond prototyping, automotive manufacturers are now increasingly bringing it into use for actual production. In recent years, additive fabrication (AM) technologies have radically changed our way to design, develop and manufacture new products. In the Automotive sector, additive fabrication technologies have made wonders to bring new shapes to life, allowing for lighter and more intricate structures at the best possible cost.

Aim and Objectives:

- Innovative and without making use of tools fabrication of intricate shape and light weight components is possible.
- Maximum freedom for design, allowing the creation of complex yet light weight components with high level of rigidity.
- Additive fabrication can significantly reduce the material waste, reduce the amount of production steps, inventory being held and reduce the amount of distinct parts needed for an assembly work.
- Reducing the need for manual assembly
- Additive makes it possible to create internal complexities and precisely control microstructure.
- Enables production of components with integrated functionality without the need for tools, thereby cutting development and production costs.
- Time and cost reduction (shorter lead time).
- Leads to more market opportunities.

Firstly we'll explain the procedure i.e., how the additive fabrication is carried out and then requirements in Automotive sector, AM in the Automotive sector, common automotive applications, current and future uses of AM in Automotive sector, challenges for Automotive sector. But before that you should have the knowledge of different AM processes.

2. Additive fabrication Procedure

I. CAD: Producing a digital model is the first step in the additive fabrication process. The most common method for producing a digital model is computer-aided design (CAD). There are a large range of free and professional CAD programs that are compatible with additive manufacture. Reverse engineering can also be used to generate a digital model via 3D scanning.

II. STL conversion and file manipulation: A critical stage in the additive fabrication process that varies from traditional fabrication methodology is the requirement to convert a CAD model into an STL (stereolithography) file. STL uses triangles (polygons) to describe the surfaces of an object. A guide on how to convert a CAD model to an STL file can be found here. There are several model limitations that should be considered before converting a model to an STL file including physical size, watertightness and polygon count. Once a STL file has been generated the file is imported into a slicer program. This program takes the STL file and converts it into G-code. G-code is a numerical control (NC) programming language. It is used in computer-aided fabrication (CAM) to control automated machine tools (including CNC machines and 3D printers). The slicer program also allows the designer to customize the build parameters including support, layer height, and part orientation.

III. Printing: 3D printing machines often comprise of many small and intricate parts so correct maintenance and calibration is critical to producing accurate prints. At this stage, the print material is also loaded into the printer. The raw materials used in additive fabrication often have a limited shelf life and require careful handling. While some processes offer the ability to recycle excess build material, repeated reuse can result in a reduction in material properties if not replaced regularly. Most additive fabrication machines do not need to be monitored after the print has begun. The machine will follow an automated process and issues generally only arise when the machine runs out of material or there is an error in the software**.**

IV. Removal of prints: For some additive fabrication technologies removal of the print is as simple as separating the printed part from the build platform. For other more industrial 3D printing methods the removal of a print is a highly technical process involving precise extraction of the print while it is still encased in the build material or attached to the build plate. These methods require complicated removal procedures and highly skilled machine operators along with safety equipment and controlled environments.

V. Post processing: Post processing procedures again vary by printer technology. SLA requires a component to cure under UV before handling, metal parts often need to be stress relieved in an oven while FDM parts can be handled right away. For technologies that utilize support, this is also removed at the post-processing stage. Most 3D printing materials are able to be sanded and other post-processing techniques including tumbling, high-pressure air cleaning, polishing, and coloring are implemented to prepare a print for end use.

3. Requirements in Automotive sector

Weight - final parts: One of the most critical aspects relating to the Automotive sector is the weight reduction of components. Automotive applications make use of advanced engineering materials and complex geometries in an attempt to reduce weight and improve performance. AM is capable of producing parts from many of the lightweight polymers and metals that are common in the Automotive sector.

Complex geometries - prototypes and final parts: Affecting weight and aerodynamics (and therefore vehicle performance) is the geometry of a part. Automotive parts often require internal channels for conformal cooling, hidden features, thin walls, fine meshes and complex curved surfaces. AM allows for the manufacture of highly complex structures which can still be extremely light and stable. It provides a high degree of design freedom, the optimization, and integration of functional features, the manufacture of small batch sizes at reasonable unit costs and a high degree of product customization even in serial production.

Temperature - testing and final parts: Many automotive applications require significant heat deflection minimums. There are several AM processes that offer materials that withstand temperatures well above the average 105°C sustained engine compartment temperatures. SLS nylon, as well as some photo-cured polymers, are suitable for high-temperature applications.

Moisture - testing and final parts: Most components that go into the production of automobiles must be moisture resistant, if not moisture proof, entirely. One major benefit of additive fabrication is that all printed parts can be post-processed in order to create a watertight and moisture resistant barrier. Additionally, many materials, by their very nature, are suited for humidity and moisture plagued environments.

Part consolidation - prototyping and final parts: The number of items in an assembly can be reduced by redesigning as a single complex component. Part consolidation is a significant factor when considering how AM can benefit the reduction of material usage, thereby reducing weight and in the long run, cost. Part consolidation also reduces inventory and means that assemblies can be replaced with a single part should repairs or maintenance need to take place; another important consideration for the Automotive sector.

3. AM in the Automotive sector

Communication: Designs in the Automotive sector often begin as scale models showcasing the form of a vehicle. These are often also regularly used for aerodynamic testing. SLA and material jetting are used to produce high detail, smooth, scale models of automotive designs. Accurate models allow design intention to be clearly communicated and showcase the overall form of a concept.

Validation: Prototyping using AM is now commonplace in the Automotive sector. From a full size wing mirror printed quickly with low cost FDM to a high detail, full color dashboard, there is an AM technology suited to every prototyping need. Some AM engineering materials also allow for full testing and validation of prototype performance.

Pre-production: One of the areas AM has been most disruptive is the production of low cost rapid tooling for injection moulding, thermoforming and jig and fixtures. Within the Automotive sector this allows for tooling to be quickly manufactured at a low cost and then used to produce low to medium runs of parts. This validation mitigates the risk when investing in high cost tooling at the production stage.

Production: Since production volumes in the Automotive sector are generally very high (greater than 100,000 parts per year) AM has predominantly been used as a prototyping solution rather than for end part fabrication. Improvements in the size of industrial printers, the speed they are able to print at and the materials that are available mean that AM is now a viable option for many medium-sized production runs, particularly for higher-end automobile manufacturers that restrict production numbers to far fewer than the average.
With the possibility of producing multiple design iterations in a shorter amount of time (and at little additional cost), 3D printing is an effective tool for product development. Typically, a part must go through several design cycles before the final design is agreed upon. With 3D printing, this stage can be sped up dramatically. Additionally, cost-efficient design improvements can be made relatively quickly, since the technology does not require expensive tooling to produce a prototype.

Tooling: Tooling is used extensively within the Automotive sector to help produce high-quality products. Additive fabrication can complement this process by creating jigs, fixtures and other customized tooling equipment. The Volkswagen Europa assembly plant is already using AM to produce tooling equipment in-house, rather than sourcing tooling equipment from third party providers. With a 10-day turnaround for positioning and screw assembly (down from 56 days using a third party source), AM clearly shows itself to be a cost-effective return on investment for tooling production, enhancing the overall production process.[3,4]

End-Part Production: Although additive fabrication was originally adopted as a prototyping tool, recent advances in AM technology and materials make the production of small and medium-size production of end parts possible. This can range from exterior components to inner parts such as bellows, complex ducting, mounting brackets, and engine components. One example is Bugatti: only this year, the luxury car manufacturer announced it had produced a fully functional titanium brake caliper — entirely 3D printed. With such breakthroughs in end-part production, 3D printing is set to become a key technology for this application.

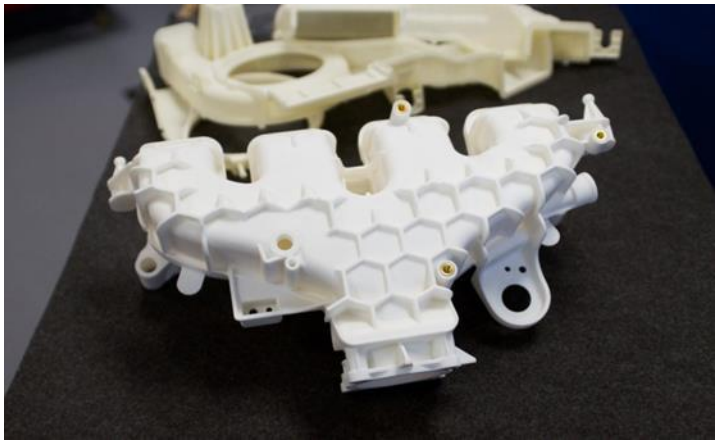
Bugattis 3D-printed brake caliper

Spare parts: Automakers can leverage the benefits of 3D printing to create spare parts on demand. With high inventory costs associated with storing spare (and often rarely ordered) parts, 3D printing provides a cost-effective means to produce parts needed on time and on demand, leading to improved delivery times, reduced inventory costs and a simplified supply chain.

4. Common Automotive Applications

Bellows: AM can be used to make semi-functional bellow pieces where some flexibility is required in assembly or mating. Generally, this material/process is best to consider for applications where the part will be exposed to very few repetitive flexing motions. For projects that require significant flexing, other Polyethylene based SLS materials such as Duraform "Flex" are better suited.

Complex Ducting: By using SLS to manufacture non-structural low volume ducting such as environmental control system (ECS) ducting for aerospace and performance racing, you can design highly optimized, very complex single piece structures**.**



A complex, functional ducting design printed in SLS nylon

With SLS it is possible to engineer in variable wall thicknesses and increase the strength to weight ratio through the application of structurally optimized surface webbing. This is a very costly detail to apply with traditional fabrication techniques. For SLS there is no cost for complexity, parts are printed without support and to a high level of accuracy.

High detail visual prototypes: Unlike traditional prototyping methods some AM processes are able to produce multicolor designs with a surface finish comparable to injection moulding. These models allow designers to get a greater understanding of the form and fit of a part. This highly accurate method of prototyping is also ideal for aerodynamic testing and analysis as the surface finish that is able to achieve is often representative of a final

part. AM is used regularly to manufacture automotive components that rely on aesthetics over function resulting in everything from wing mirrors and light housings to steering wheels and full interior dashboard designs being produced. Material jetting and SLA printing are the two most common methods used for aesthetic prototypes producing parts from a photo-activated resin.

Full colour, textured visual prototypes like this centre console can be produced via AM

Functional mounting brackets: Being able to rapidly manufacture a complex, lightweight bracket overnight is a trademark of the AM industry. Not only does AM allow for organic shapes and designs to be manufactured but AM also requires very little input from an operator meaning that engineers are able to quickly take a design from a computer to assembly in a very short amount of time. This is not possible with traditional fabrication techniques like CNC machining where a highly skilled machine operator is needed to produce parts. Powder bed fusion technologies like SLS nylon and metal printing are best suited for functional parts and offer a range of materials (from PA12 nylon to titanium).

A functional alternator bracket printed using SLS nylon

5. Current Uses of Additive fabrication

Exhausts and emissions:
melting to create cooling vents.

Fluid handling: Selective laser melting and electron beam melting are utilised with aluminium alloys. These techniques can be used to make pumps and valves within the fluid handling system.

Exterior: Using selective laser sintering, polymers are currently used to manufacture wind breakers and bumpers.

fabrication process: Hot work steels and polymers can be used together with a variety of additive fabrication processes such as selective laser sintering, selective laser melting and fused deposition modelling for prototypes, casting and customised tooling.
The above uses in the Automotive sector are used from both small companies and large international conglomerates.

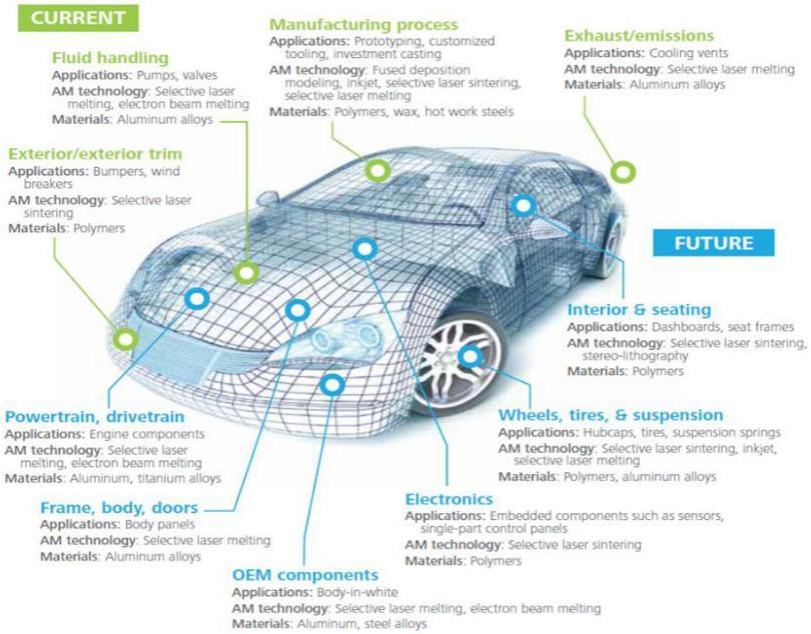
Interior and seating: Using polymers and the techniques of stereo-lithography and selective Laser sintering, dashboards and seat frames could be manufactured.

Tyres, wheels and suspension: Aluminium alloys and polymers can be manipulated with the aid of selective Laser sintering, selective laser melting and inkjet technology to create suspension springs, tyres and hubcaps.

Electronics: Selective laser sintering can be used on polymers to manufacture a range of delicate components including parts which have to be embedded, such as sensors, and single part control panels.

Framework and doors: Selective laser melting can be used on metal compounds such as aluminium alloys to create body panels, including framework and doors.

Engine components: Various functional parts of the engine can be made from metals such titanium and aluminium allows when techniques such as electron beam melting and selective Laser melting are used. [1,2]



6. Challenges for the Automotive sector

Mass production: While automotive OEMs are increasingly incorporating AM systems into development and production, one challenge to wider adoption is a production volume. With over 80 million cars produced in 2017 alone, the Automotive sector is heavily reliant on mass, series production. 3D printing should therefore not be seen as a replacement for traditional fabrication methods, well-suited to mass volumes, but as a complementary tool for lower volume, customised end parts.

Build sizes: Another challenge faced by automakers is the limited build size of many AM systems. Although larger parts can be produced with 3D printing technology, this must be done in the form of modular parts. These in turn currently have to be assembled or attached together through other processes, such as welding. However, large-scale additive fabrication is an important and growing area of research, with technologies that can support larger build sizes, such as Wire Arc Additive fabrication (WAAM) and Big Area Additive fabrication (BAAM), actively being researched and developed.

AM skills gap: Additional investment into developing AM-specific skills must also be addressed if the technology is to take off more widely. Design for additive fabrication as well as the operation and maintenance of AM systems, materials and post-processing are all vital skills that must be developed and nurtured. While much has been said on the current skills gap for AM, partnerships with universities and internal training programmes are one way of ensuring a skilled talent pool able to work with the particularities of AM technology.

Conclusion

This paper aimed to figure out the scenario of additive fabrication in Automotive sector. Improvements in additive fabrication materials and product quality are expanding the usefulness of 3D printing in auto fabrication. Some parts produced using AM technologies occasionally have tiny “voids” or pores that can weaken overall strength. In others, dimensional accuracy is not always on par with parts made with conventional fabrication processes. These and other quality issues can diminish product uniformity and consistency, a challenge for high-volume industries such as automotive in which quality and reliability are critical.

Profitability in the Automotive sector is driven by volume. Given these enormous volumes, the low production speed of AM is a significant impediment to its wider adoption for direct part fabrication. This has made high-speed AM an important area of research.

Despite of this, AM offers a versatile set of technologies that can support auto companies as they pursue performance, growth, and innovation. Traditional fabrication techniques are deeply entrenched and will likely hold a dominant position in the Automotive sector for the foreseeable future. Yet the breadth of AM capabilities—and the success of on-going efforts to broaden their application—suggest that going forward, additive fabrication will also play an important role in shaping the global automotive landscape.

Although, the doors for conventional fabrication are still open and will play a dominant role in automotive fabrication, additive fabrication is making inroads and is obvious to change the global shape of the industry.

References

[1] D. Cooper, J. Thornby, N. Blundell, R. Henrys, M. A. Williams, and G. Gibbons, "Design and Manufacture of High Performance Hollow Engine Valves by Additive Layer fabrication,"Materials & Design, vol. 69, pp. 44-55, 2015.

[2] S. Biamino, B. Klöden, T. Weißgärber, B. Kieback, and U. Ackelid, "Properties of a TiAl Turbocharger Wheel Produced by Electron Beam Melting," in Fraunhofer Direct Digital fabrication Conference, Berlin, Germany, 2014.

[3] M. Cotteleer, M. Neier, and J. Crane, "3D Opportunity in Tooling: Additive fabrication Shapes the Future," Deloitte University Press 2014.

[4] A. V. Villalon, "Electron Beam Fabrication of Injection Mold Tooling with Conformal Cooling Channels," M.Sc. Thesis Industrial Engineering North Carolina State University, Raleigh, North Carolina, USA 2005.

[5] https://www.sculpteo.com/blog/2016/01/20/3d-printing-transforms-the-automotive-industry/

Design and Analysis of Hollow and Solid Shaft

GOVINDOLA VEERANDER GOUD

veerander85@gmail.com

M Tech Student, Department of MECH,

Malla Reddy college of Engineering, Hyderabad, India.

Abstract

The Roadway vehicles like vehicle, buses, trucks and land movers goods many mechanic ability in common like Engine parts, Propeller shafts, Gearbox, Brakes, Clutches, Wheels, etc., To make the vehicle kindling capable which in result mate the transportation economical, the weight of that vehicle should be reduced. Since the composite materials are publicity weight with more puissance & hardness, inclusion of composite materials to conventional steel materials custom in auto parts will lessen the weight and better the machinelike properties of those components. In this thesis, deals with shaft of MARUTHI OMNI to design the shaft for its minimum dimensions to suffice authentic question specification and then replace accepted steel material with composite material. The design of the propeller shaft is first

theoretically designed for steel, aluminum allay ,lose iron and kelvar composite essential for its safe dimensions. Then they can be created as a part shape for respective dimensions in CATIA software. After modeling, static analysis and Modal analysis can be carried out in the propeller shafts worn

Keywords: **Shaft; Coupling; ANSYS; CATIA;**

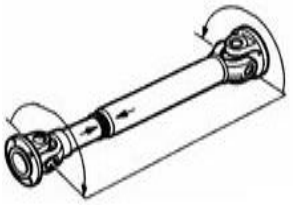
1. INTRODUCTION

SHAFT: Propeller Shaft is the shaft that transmits Command from the gear case to the differential gear in a motor vehicle from the engine to the propeller in a sail or flying machine.

Fig 1.1.1 Propeller shaft

Propeller shaft, sometimes exhort a card an well,

- Used in vehicles with a short distance between the engine and axles, and MR based four- wheel-drive vehicles.
- The friction wielding adopted at the junction contributes to an



ANSYS software.

Fig 1.1.2 hallow shaft

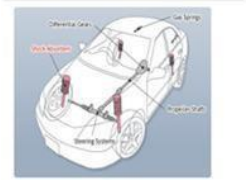
Types of Propeller Shaft: Single–

Piece–Type Propeller shaft :

Improvement in the strength,quality and durability of the junction

2-piece-type/3-piece-type

- Utilized as a part of vehicles with a long distance between the engine and axles and front engine front drive base foure-wheel-drive vehicles
- The division of the propeller shaft into two- or three- parts allows the critical number of revolutions to lowered preventing vibration issue from occurring, when the overall length of the shaft increased



- A tubular-section propeller shaft is normally used as it has (i) Hill weight, (ii)provides low resistance to misalignment, especially sag, (Hi) has serviceable torsional Strength , and (iv) provides low resistance (low inertia) to exchange in sharp quickness, which arise when a hookes type coupling is used to drift the shaft. Since a propeller shaft often revolve at high speed,

2. METHODOLOGY

I. Material modeling for the properties of nano composite, this has been Modeled using CATIA parametric software

ii. Modeling of propeller Shaft Model by using CATIA parametric software.

iii. Determination of stress, strain, deformation and frequencies.

iv. Modal analysis of the propeller shaft model.

3. LITERATURE REVIEW

Functions: Propeller shaft is a automatic constituting for transmitting torque and rotation, usually custom to connect other components of a drive discipline that cannot be connected expressly long of distance or the need to allow for relative movement between them. Propeller shaft make any machine move. Drive well is carriers of torque: they are subject to torsion and fleece urge, analogous to the difference between the input torque and the freight. They must therefore be strong enough to bear the stress, whilst avoiding too much additional load as that would in turn increase their inertia. The longer the shaft, the more bound it is to incline and bending is further promoted when rotation is address reason vibrations and resulting in an augment in noise. For this ground, the propeller shaft has been show to overwhelm vibrations proceed from a remote range of source.
Propeller Shaft Vibration: Small auto and short vans and trucks embodied a single propeller well with a omit-joint at the front end without having any undesired vibration. Vehicles with longer wheelbase need longer propeller shaft, which has a disposition to sag and to harry under certain operating circumstances (Fig. 26.3). As a result resonant vibrations are put up in the body of the vehicle, so that the body vibrates as the shaft whirls. The might agent responsible for the rebellant frequency of the propeller shaft causing the vibration may be grouped as syn Propeller Shaft Intermediate Support Bearings: Intermediate bearing-and-mount assemblies are incorporated to position and support the divided propeller shafts. These assemblies are either of (i) self-aligning bearing verify represent of («) obsequious-mounted influence assist semblance. Self-aligning intermediate-bearing supports are mostly employment on heavy-respect trucks. One type of this bearing support is a double-line courage-conduct with a deep-fossulate inner race and an internally semicircular external race (Fig. 26.7A). This arrangement compensates any shaft inflection through the inner line and nonsense, which tilt about the fixed dispossession-race sphere-shaped post.

4. RELEATED STUDY

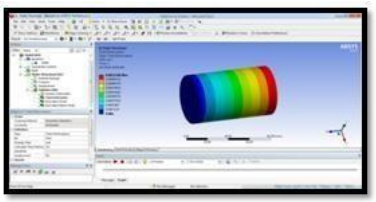
INTRODUCTION TO CATIA: (an acronym of computer aided three-dimensional interactive application) is a multi-plan software suite for electronic computer-aided purpose (CAD), computer-assist manufacturing (CAM), computer-relieve engineering (CAE), PLM and 3D, developed by the French party Dassault Systems. CATIA started as an in-tenement unraveling in 1977 by French aircraft manufacturer Avions Marcel Dassault, at that measure purchaser of the CADAM software to develop Dassault's Mirage fighter jet. It was later adopted by the aerospace, self-propelled, shipbuilding, and other industries. Initially denominate CATI (conception assisted tridimensionally interactive – French for interactive aided three-dimensional project), it was renamed CATIA in 1981 when Dassault appoint a assisting to develop and sell the software and type a non-exclusive distribution agreement with IBM.

HALLO SHAFT

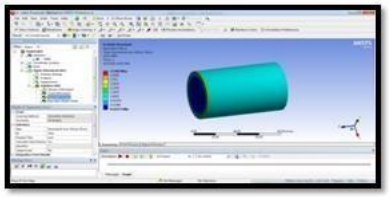
SOLID SHAFT

ANSYS Software: ANSYS is an Engineering Simulation Software (computer relieve Engineering). Its bowl shelter Thermal, Static, Dynamic, and Fatigue finite element analysis along with other use all designed to help with the development of the fruit. The party was founded in 1970 by Dr. John A. Swanson as Swanson Analysis Systems, Inc. SASI. Its primary view was to develop and nundinal finite element analysis software for stextural physics that could simulate static (motionless), functioning (drifting) and ardor transfer (thermal column) problems. SASI developed its business in analogue with the growth in computer technology and engineering needs.

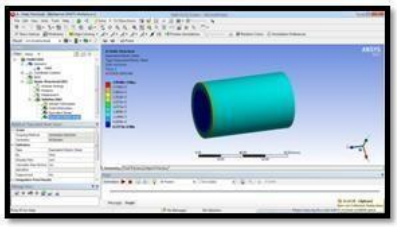
Stress



Strain



Total deformation

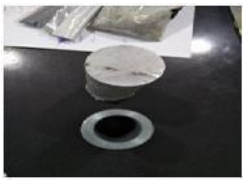


5. RESULT TABLES

STATIC ANALYSIS RESULTS

Models	Material	Deformation (mm)	Stress (N/mm^2)	Strain
Solid	Aluminum	0.0011106	2.1263	2.9948 e-5
	Aluminum with BLF	0.00011475	2.0117	2.9193 e-5
Hollow	Aluminum	0.0012224	1.7055	2.639e -5
	Aluminum with BLF	0.0013249	1.6791	2.7984 e-5

PRACTICLE EXPERIMENT



6. CONCLUSION

The aluminum with BLF composite solid and hollw shafts are intend to meet safe design requirements as the stipulated steel shaft. From the static analysis the deformation, ,VonMises distress and weight are determined. In overall similitude aluminum with BLF composite hollw pit is correct only in weight curtailment and that too only 1.16% lesser weight than aluminum fineness with BLF compounded shaft.

7. REFERENCES

[1] Branislav Duleba, Ludmila Dulebova and Emil Spisak (2014), 'Simulation and evaluation of carbon/ epoxy compound system second-hand FEM and tensile test', Science Direct – Procedia Engineering 96(2014), pp.70-74.

[2] Srinivasa Moorthy R, Yonas Mitiku and Sridhar K (2013), 'Design of Automobile Driveshaft using Carbon/Epoxy and Kevlar/Epoxy Composites', American Journal of Engineering Reasearch (AJER) – Vol.-02, Issue-10, pp-173-179.

[3] Vicente Sanchez-Galvez, Laura Sanchez Paradela and Francisco Galvez (2014), 'Analytical feint of dear speed impact onto lurcher glass/carbon epoxy composites targets', Science Direct – Procedia Engineering 88(2014), pp.101-108.

[4] Bhushan K Suryawanshi and Prajitsen G Damle (2013), 'Review of Design of Hybrid Aluminum/Composite Drive shaft for Automobile', International Journal of Innovative Technology and Exploring Engineering (IJITEE) – Vol.-2, Issue-4, pp-259-266.

[5] Dr Andrew Pollard (1989), 'Polymer Matrix Composites in Driveline Applications', Journal of Composite Structures, Vol.25, pp- 165-175.

[6] Agarwal B D and Broutman L J (1990), 'Analysis and deed of fiber composites', John Wiley and Sons Inc.

[7] Rajanish M, Nanjundaradhaya N V and Ramesh S Sharma (2014), 'An investigation on ILSS Properties of Unidiresctional; Glass Fibre Alumina Nanoparticles Filled Epoxy Nanocomposite at Different angles of Fibre Orientations', Science Direct-Procedia Materials Science 10(2015), pp.555-562.

[8] Palanisamy P, Sureshkumar D and Balamurali S (2014), 'Hybrid Design and Analysis of Advanced Propeller shaft', International Journal of Engineering Sciences and Research Technology – Vol.-3, Issue-5, pp.123-127.

[9] Murugan R, Ramesh R and Padmanabhan K (2014), 'Investigation on Static and Dynamic Mechanical Properties of Epoxy based Woven fabric Glass/Carbon Hybrid Composite Laminates', Science Direct-Procedia Engineering 97(2014), pp.459-468.

SUPERIOR EQUIPMENT AND VARNISH OUTLOOK FOR GAS TURBINE APPLICATIONS

B.SAI VENKATA KRISHNA[1], Asst.Prof, Department of Mechanical Engineering.
E-mail: sai331988@gmail.com , Megha Institute of Engineering and Technology for Women, Hyderabad.

M.A.AZEEM[2], Asst.Prof, Department of Civil Engineering.
E-mail: azeem26jan@gmail.com , Megha Institute of Engineering and Technology for Women,Hyderabad

ABSTRACT:

The need for very specific hardness / hardness materials can only be satisfied by design concepts using reinforced compounds. Carbon fiber made of high-strength and durable silicon in a high-temperature titanium matrix is one of the main candidates whose development will be described. Design concepts for high-pressure turbines that incorporate thermal protection layers for ceramics, i.e. heat-insulating coatings, will go beyond the natural limits provided by the melting point of Ni-based super-turbine blades. The sophisticated design of the aerodynamic engine will focus on reducing specific fuel consumption and increasing the weight-to-weight ratio. In the end, this requires an increase in pressure ratios, as well as higher operating temperatures, and certainly poses a major challenge to the structural design and materials used. High-capacity materials for high temperatures are required, as are very light structures. Reducing the weight of the pneumatic dynamic motor requires a new compact compressor design with a few phases. Gas turbine blades are designed for cooling methods, as well as for cooling films in external cooling and thermal cooling in internal cooling. The turbine blade is designed with four-hole and six-hole heat cooling. The film cools the air in the blade through several small holes in the chassis. The current material used for the blade is chrome-plated steel. Here, it is replaced by composite materials of ceramic matrix and silicon carbide. Advanced evaporation treatment using electron beam technology is the preferred choice for manufacturing these coatings in highly rotating parts. However, considerable efforts are still needed to improve these coatings, make them more reliable, and thus achieve a philosophy designed to fully exploit their potential.

Keywords: *gas turbine, rotor blade, steady state thermal analysis, fiber-reinforced composites, thermal barrier coating.*

1. INTRODUCTION

During the design of the new aero engine, consideration should be given to the impact on key engine characteristics such as engine weight, specified fuel consumption (SFC), manufacturing costs and serviceability. For airlines, the direct operating costs of the aircraft are an initial qualifying parameter. Regardless of the specific fuel consumption, the propulsion / weight ratio is of great importance for military engines. The pay-to-weight ratio has been significantly improved over the years achieved by increased

operating temperature and structural efficiency. It is very clear that advanced materials play a key role here. In fact, modern aircraft engines represent some of the most demanding and sophisticated building material applications in any engineering system manufactured today [1]. The first addresses recent developments in high pressure turbines / low pressure turbines and focuses on light titanium alloys, titanium aluminates and especially titanium matrix compounds (TMCs). The second covers the thermal barrier coatings (TBCs) of high pressure turbine blades that try to get rid of the bottleneck in developing improved performance engines. The excellent properties of titanium alloys include high specific strength and excellent corrosion resistance. Therefore, titanium alloys are found in aerospace applications where a combination of weight, strength, corrosion resistance and / or high temperature stability of light metal aluminum, high strength steel or nickel based super alloys is insufficient. In air motors, titanium alloys represent the most important class of engine compressor materials. The compressor blades were the first engine components made of titanium, and the titanium compressor discs are then introduced. The large front fan blades of modern jet engines are also often made of titanium alloy. Due to the constant increase in engine deviation rates, the latest blade designs exceed one meter in length. In these dimensions, the fan blade blade can become a serious problem because the blade edges can reach the speed of sound and cause acoustic / subsonic flow areas and associated shock waves. Advanced fan designs have improved blade stiffness by increasing column width and reducing the number of blades by about one third. Today, these "wide propeller blades" are used in the latest jet aircraft engines [2]. The new Benz (G 800) and GE / Pratt & Whitney Engine Alliance (GP7200) Airbus A380 engines will be about three meters in diameter and will include hollow titanium blades. Blisk technology is now standard in low and medium sized compressors for commercial and military engines. In the Eurofighter EJ200 engine, for example, the three stages of the fan section are of excellent design; the first two are manufactured by linear friction welding and the third by ECM.

Fig 1: Three stage blisk compressor

2. METHADOLOGY

This temperature limit for titanium alloys means that the hottest parts of the compressor, i.e., the discs and the blade in the later stages of the compressor, must be made of twice the weight of nickel-based super alloys. In addition, problems arise related to different thermal expansion behaviors and bonding techniques in the two alloy systems. Therefore, huge efforts are being made to develop a compressor made entirely of titanium. Titanium alloys that can be used at temperatures around 600 ° C or higher are required. This has been the motivation for intensive research and development in the field of titanium aluminize. These substances, based on intermetallic compounds α2 (Ti3Al) and γ (TiAl), have been studied for their ability to raise the application temperatures of titanium alloys to 650 ° C and 800 ° C, respectively. Excellent creep resistance is due to the organized nature of the crystalline structure. However, this structure also makes the intervals relatively fragile and difficult to distort. Alloys with Nb, Cr, V, Mn or Mo and microstructure optimization are two ways to increase ductility. Adequate tolerance for damage, pathological oxidation behavior and productivity (cost) are key factors that will determine the use of titanium acuminate in the aviation industry [3]. Due to the high reactivity of titanium alloys with SiC granules, manufacturing processes that occur with the least possible thermal load on the compound during manufacturing are preferred. Thus, processes based on vapor deposition and solid-state formability is considered. Today, the preferred route is fiber-coated matrix technology. The primary product is homogeneously coated matrix fibers allowing the manufacture of composite materials with excellent fiber coordination and precise matrix structure. Deposition of the magnetron spray layer from the vapor phase. Due to the high deposition rate, the electron beam vapor deposition (EB-PVD) is also used, but is limited to individual composition matrix alloys. In the second step, the matrix fibers are assembled or arranged using matrix, for example, winding techniques to achieve the desired geometry of components, encapsulated and then compressed at an even temperature at a constant temperature of about 950 ° C and pressures of about 2000. Tape. In the last step, the component is formed to its final geometry.

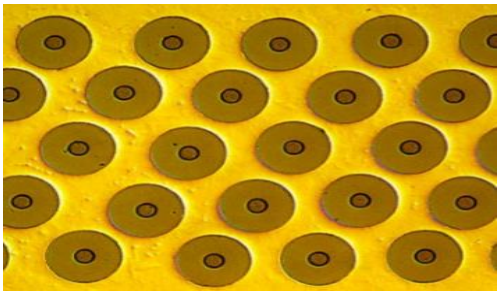


Fig 2: Titanium matrix composite

3. LITERATURE SURVEY

PAPER 1- Design and Analysis of Gas Turbine Blade by Theju V, Uday P S , PLV Gopinath Reddy , C.J.Manjunath

The purpose of this project is to design and analyze turbine pressure. Research is required to use new materials. In the present work, the turbine blade is designed with different materials, Inconel 718 and Titanium D-6. Try to check the effect of temperature and pressure caused by the turbine code. Thermal analysis was performed to verify the direction of temperature flow due to convection. Structural analysis was performed to investigate the effect of the combination of thermal loads and centrifugal effects and shear forces and turbine plate displacement. We have tried to suggest the best fit for the turbine index by comparing the results obtained from two different substances (Inconel 718 and T6 Titanium). Based on the designs and results, Inconel 718 can be considered the best economical material and has good physical properties at higher temperatures compared to titanium T6.

PAPER 2- Heat Transfer Analysis of Gas Turbine Blade Through Cooling Holes by K Hari Brahmaiah , M.Lava Kumar

In advanced gas turbines, the operating temperature of the turbocharger operates above the melting point of the blade material. An advanced cooling system should be developed to ensure the safe and continuous operation of high performance gas turbines. Different ways of cooling the blades are offered. One of these techniques is the presence of radial holes at high air cooling speeds along the plate. Heat transfer analysis of gas turbines was performed on four different sample blades without holes (5, 9 and 13) with holes. Reinforced wall). For heat transfer graphics and overall distribution, the code is optimized for 13 holes. A consistent and consistent analysis is performed using ANSYS with different chrome-plated steel blades and Inconel 718. Comparison of these materials shows that although Inconel 718 has excellent thermal properties, the induced stress is lower than chromium steels.

PAPER 3 - Film Cooling of the Gas Turbine Endwall by Discrete-Hole Injection by M. Y. Jabbari, K. C. Marston, E. R. G. Eckert and R. J. Goldstein

The film's cooling performance is tested for injection through separate openings at the end of the turbine blade. Efficiency is measured at about 60 sites in the injection area. Three nominal hit rates, two density rates and two Reynolds flow numbers are examined. Data analysis reveals that up to 60 locations are insufficient to determine the extent of the film's cooling effect with its strong local variations. Viewing the effects of cooling jets on the perimeter wall using ammonium diazo paper provides useful quality information for interpreting measurements, pathways and jet responses that change at fast speed and flow rate. intensity.

4. RELEATED STUDY

In general, composite materials are highly dependent on the properties of the solvent. Understanding the interaction between them forms the basis for physical evolution. The specificity of embedded properties always depends on the stability of these reactions. Long-term durability and robustness to the highest service temperatures are definitely a highlight of the enhanced TMCs. Energy, or more energy, energy associated with mass, is an important factor for structural simplicity and is a valuable property of great importance in design [4]. Dynamic aerospace and TMC data are shown in Figure 6 for temperatures up to 700 ° C. Although they are strong (or medium), the aluminum alloys, α2Ti3Al or mushrooms (Ti2AlNb), al-alumate, and suparalloysi (here 718) cover 10 to 25 km at room temperature and from 10 to 15 km at 800 ° C, The maximum TMC strength ranges from 40 to 60 km at room temperature and up to 50 km at 800 ° C. Maximum strength depends on the fiber part. A maximum of 40% electrical fracture was obtained. For SiC / TIMETAL 834, the maximum heat resistance of 2400 MPa is recoverable, which is well consistent with the calculations according to chemical law. The use of TMCs in advanced applications for advanced jet engines requires less energy; if the behavior is weak under heavy load, it is one of the most important criteria. Figure 7 shows the nonlinear force with reinforced SiC fibers 5 Christoph Leyens TIMETAL 834 under friction jump at temperatures up to 700 ° C. The TMCs have a positive effect, particularly at 700 ° C where the fiber strength is high in conductivity. The periodic pressure levels of TMC antagonists in LCF (low fatigue state) and HCF (high fatigue condition) were more than 100% greater than nonlinear ones. Although the maximum continuous closure is about 400 MPa, SiC / Ti reaches a fixed limit of more than 1000 MPa at 700 ° C.

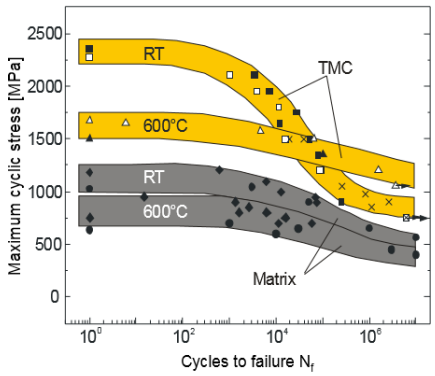


Fig. 3: Fatigue strength of TMCs in comparison

5. APPLICATIONS

The future development of natural gas is clearly aimed at raising the temperature of the turbine marine to more than 1700 ° C. There is no doubt that this ambitious goal can be achieved through by using non-economical televisions or by high temperature applications, in particular by expanding the use of heat exchangers (EB-PVD) using electromagnetic fields (TBCs).). Chlorofluorocarbons (CFCs) consisting of hot-tempered layers with reduced thermal conductivity - usually zirconia-stabilized particles - are used in the air and only in powder form skin. The paint showed good consistency of cement in the material [5] [6]. The TBCs application allows for increased engine / performance by increasing the temperature or reducing air conditioning. On the other hand, it is possible to extend the life of turtles by reducing the heat of iron as shown. During the process of EB-PVD, the strong electric field is melted and the cement material is transferred into a vacuum chamber. The cells are inserted internally during the healing process to ensure continued TBC. To achieve the measurement of specific elements of zirconium, the amount of oxygen is compromised in the discharge chamber. Intravenous reservoirs are placed on a storage tank at a rate of 3 to 30 microns / min. Specialized and polished microspheres provide smooth surfaces without the need for final drying or cooling. Due to the microscopic structure, the life of TBCs is prolonged and impairs tolerance. Characteristics and many advantages of TBCs compared to heat exchanger and evaporator featuring EB-PVD TBC on a steam engine powered by DLR using 150kW von Ardenne EB-PVD.

6. CONCLUSION

Mechanical devices, including high-strength, high-strength and high-strength fiber-reinforced titanium matrix materials, are ideal for high-tech applications, for example in turbine engines. Due to high costs and lack of knowledge of real estate, their use is severely restricted by applications. The following procedures should enhance TMC's ability to increase acceptance in the manufacturing market. Although most TMC applications today focus on temperature, the future of TMC is likely to be a future at high temperatures. For large-scale heat exchangers such as power plants, Eb-PVD shows the highest potential for increasing turbine power. TBCs represent a complex process involving (at least) attachments, bandages, heat transfer heaters and top cover. Each species can affect the life cycle of tuberculosis through a strong immune system. Complex and commonly used conditions, thermal, mechanical and mechanical conditions of the equipment, including heat transfer and cyclic effect, aggravate the situation.

7. REFERENCES

[1] J.C. Williams, The Development of Advanced Gas Turbines: the Technical and Economic Environment, Proc. Materials for Advanced Power Engineering 1994, ed. D. Coutsouradis, et al., Kluwer Academic Publishers (1994) 1831-1846

[2] M. Peters, J. Kumpfert, C.H. Ward, C. Leyens, Titanium Alloys for Aerospace Applications, in: Titanium and Titanium Alloys, Advanced Engineering Materials 5 (2003) 419-427

[3] F. Appel, M. Oehring, Gamma-Titanium Aluminide Alloys: Alloy Design and Properties, in: Titanium and Titanium Alloys, C. Leyens, M. Peters (eds.), 89152, Wiley-VCH, Weinheim, Germany, 2003

[4] H. Kestler, H. Clemens, Production, Processing and Applications of γ(TiAl)-Based Alloys, in: Titanium and Titanium Alloys, C. Leyens, M. Peters (eds.), 351-392, Wiley-VCH, Weinheim, Germany, 2003

[5] R. Leucht, H.J. Dudek, Properties of SiC-fibre reinforced titanium alloys processed by fibre coating and hot isostatic pressing, Materials Science & Engineering A188 (1994) 201-210

[6] C. Leyens, M. Peters, W.A. Kaysser, Oxidation Resistant Coatings for Application on High Temperature Titanium Alloys in Aeroengines, Advanced Engineering Materials 2 (2000) 1-5

THERMAL ANALYSIS OF MODIFIED COMBUSTION CHAMBER OF SPARK IGNITION ENGINE

BANOTH RAJENDRA PRASAD
MallaReddy College of Engineering, Masimmaguda, Hyderabad 500014, Telangana state

N. KISHORE KUMAR
MallaReddy College of Engineering, Masimmaguda, Hyderabad 500014, Telangana state

V.SRINU
MallaReddy college of Engineering, Masimmaguda, Hyderabad 500014, Telangana state

ABSTRACT

Efficiency of an internal combustion engine can be increased by increasing the heat energy generation inside the combustion chamber without effecting the performance of lubricating oil and lessening the strength of the piston. The main aim of this paper is to determine temperature distribution in a four-stroke, single-cylinder, water cooled, variable compression ratio (3-9), variable speed (2200-3000 rpm) spark ignition engine with brake power of 2.2 kW at a speed of 3000 rpm with copper coated combustion chamber (CCE) [copper-(thickness, 300 μ) was coated on piston crown, inner side of liner and cylinder head] and compared the engine, with conventional combustion chamber (CE) with neat gasoline operation. Copper coated power piston, copp.

Keywords: Copper Coating, Piston, Liner, Solid Works, Ansys Workbench, Thermal Analysis.

1. INTRODUCTION

With advent of urbanization, energy consumption is increasing drastically, out of which gasoline energy utilization is foremost as transportation development is fast. Though alcohols are important substitute for gasoline to overcome the fossil fuel crisis, it plays the major contribution for the exhaust pollutants in SI engine due to incomplete combustion of fuel [1]

Engine modification with copper coating on crown of the piston and inner side of cylinder head improves the engine performance as the copper is a good conductor of heat, stabilizes flame, improves pre–flame reactions and turbulence [2-5]. Temperatures inside the combustion chamber are very high and the heat generated is transferred to different components of power piston ,liner and cylinder head such that the need for the study of temperature distribution across the components is a must, especially to provide proper lubrication between liner and power piston and cooling to the walls of combustion temperature [6-10]. Finite Element Analysis is applied in such situations and thereby temperature distribution across the components is determined by using ANSYS programme which employs finite element-based software

1. MATERIALS AND METHODS

1.1 Making of copper coated combustion chamber

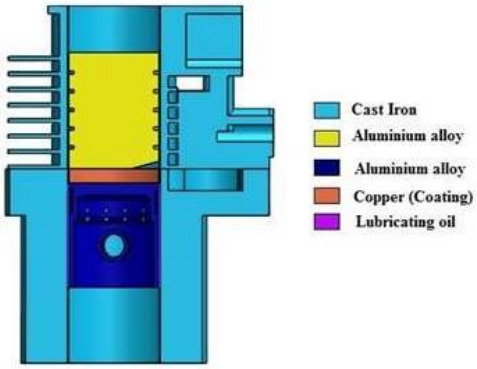
In the copper coated engine, top surface of the power piston, top surface of auxiliary piston an inside of liner are coated with copper using Twin wire spray gun. For 100μ thickness, nickel-cobalt-chromium bond coating was sprayed. On this coating, for another 300μ thickness, an alloy of copper (89.5%), aluminum (9.5%) and iron (1%) was coated with a MEC (Trade name of the

company) flame spray gun. The bond strength of the coating was so high that it does not wear off even after operating it for 50hrs continuously.

The specifications of the engine are given below

ENGINE SPECIFICATIONS

1 Type: Four-Stroke, Single Cylinder, Variable
2 Make: Greaves Limited
3 Rated Power:2.2 KW at 3000rpm



4 Bore and stroke:70mm x 66.7mm
5 Speed:2200rpm-3000rpm
6 Compression Ratio: 3:1 to 9:1
7 Spark Plug: Make: MICOBOSCH
8 Spark Plug Gap:0.6mm
9 Type of Ignition: Battery
10 Specific fuel consumption:500 gm/h KW
11 Lubricating Oil: SAE-40
12 Dynamometer: Eddy Current Dynamometer loading Rheostat
13 Temperature and Speed: By Digital Indicators
14 Starting: Auto Start by DC Motor
15 Cylinder Pressure: By Sensor, range :5000PSI
16 Exhaust Gas Calorimeter: IND-LAB Make
17 Torque arm distance:200mm
18 Orifice Diameter:20mm
19 Recommended Spark Ignition Timing: 25

2.2. Measurement of Temperature distribution of engine components

In the present scenario, steady state thermal analysis is done on the assembly of four stroke variable compression spark ignition engine to calculate temperature distribution, heat fluxes, temperature gradients and amount of heat lost or gained in engine components for different versions of the engine. Steady state thermal analysis is done in two main steps. First Geometric Modelling is done then Finite Element analysis is done.

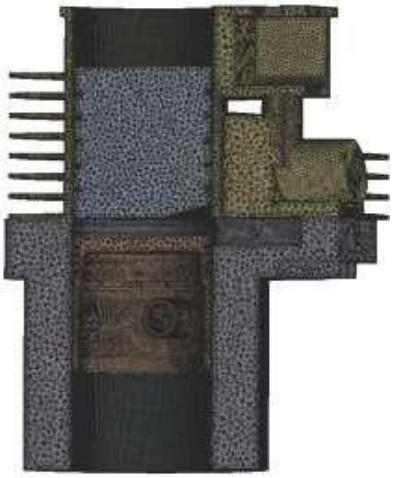
2.2.1. Solid Model Creation: In Geometric Modelling, 3-D geometry of power piston, auxiliary piston, liner and cylinder head are created. The models are generated using SOLID WORKS 16**.** The components are assembled by giving constraints to obtain the final assembly. The assembly is saved as. IGS file. The file is imported to ANSYS workbench15.

Figure 1 shows the Configuration of the assembly with cu coating piston, liner and cylinder head and showing different materials in various zones

Figure 1 Configuration of the assembly witcu coating piston, liner and cylinder head showing different materials in various zones

2.2.1. Meshing the model: In finite element modelling, mesh was generated with tet10 element type using ANSYS WORKBENCH 15. Since the geometry is complex, free style of meshing is employed. Fine mesh size is considered at small and critical components, while coarse mesh is considered for the remaining components.

Figure-2 Mesh employed in the thermal analysis for the assembly piston, liner and cylinder head with cu coating.



2.2.3 Boundary Conditions Application and Thermal Analysis Solution: The boundary conditions for the present problem are obtained from [11]

The top surface of the power piston is applied with a convective heat transfer (h_c) of 250 w/m^2 k and a bulk temperature (T) of 920 ^{0} C and on the water jacket side of the liner, $h_{c=}$ 1800 w/m^2 k and T=60 ^{0} C and on the cylinder head fins of h_c =120 w/m^{2}k, T = 60 ^{0} C.

Steady state thermal analysis was solved after application of boundary conditions

3. RESULTS AND DISCUSSION

3.1 Thermal Analysis Results

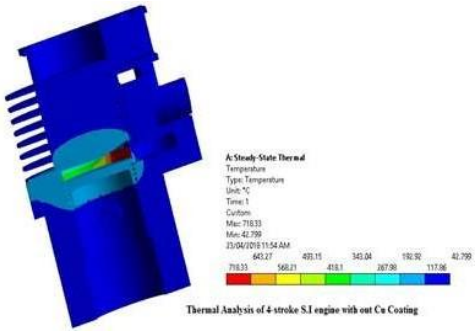


Figure 3 Isotherms of thermal analysis for the assembly of piston, liner and cylinder head of CE

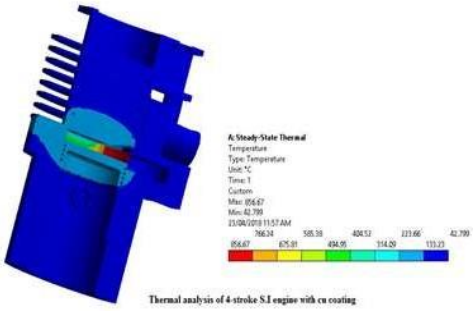
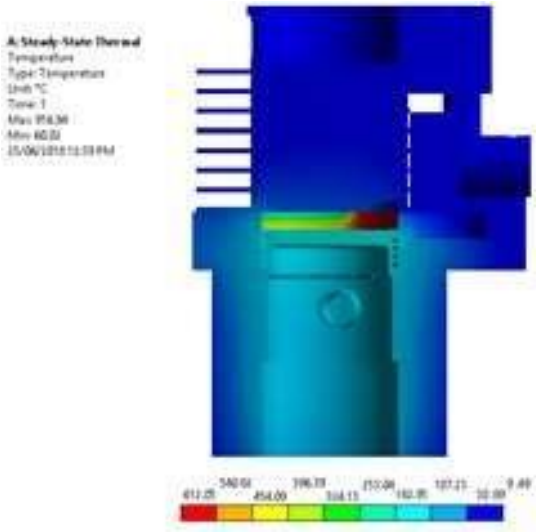


Figure 4 Isotherms of thermal analysis for the assembly of piston, liner and cylinder head of CCE

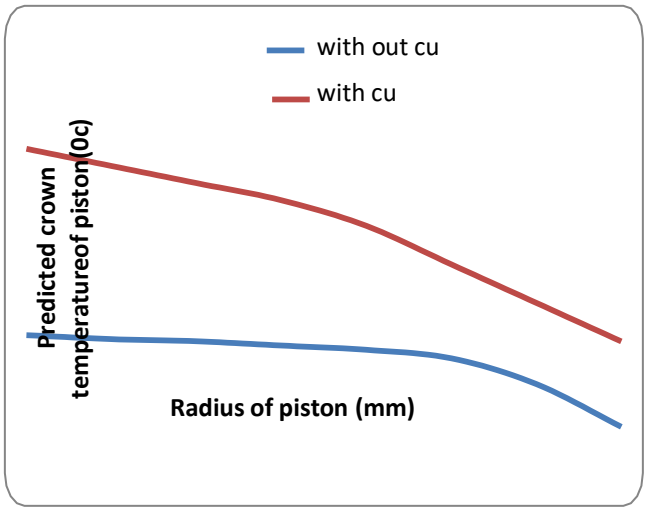


Prediction of the Temperature along the Piston Radius for the Base Engine and Catalytic Coated

Engine

The fluctuation in the crown temperature of the piston predicted by FEM analysis with radius of the piston for CE and CCE was shown in Figure-5

Figure 5 Variation of the predicted crown temperature with radius of the piston for CE and CCE.



From the Figure 5 it was noticed that, as the radius of piston increases, crown temperature of the piston decreases marginally for both CE and CCE. The temperature decreases at the outer periphery of the piston, as it is cooled by means of lubricating oil and also with the presence of fins. The temperature of the piston of CE is 181^0C while it is 225^0C for the piston of CCE at the crown surface. At the outer periphery of the piston, the temperature decreases to160^0C for CE and 180^0C for CCE. The temperature drop for the piston of CCE from the crown to the outer periphery was less as the piston is coated with copper, which has high thermal conductivity and hence thermal resistance was less.

Figure 6 Heat flux in the assembly of piston, liner and cylinder head of CE

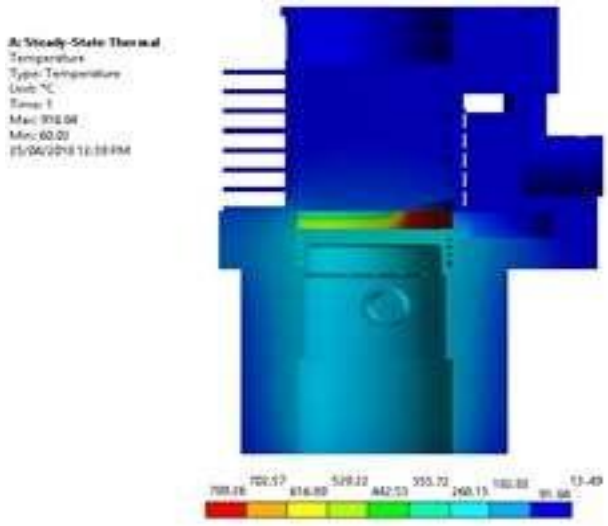


Figure 7 Heat flux in the assembly of piston, liner and cylinder head of CCE

Prediction of the Heat Flux along the Radius of the Piston in Base Engine and Catalytic Coated Engine

The variation of the percentage (%) increase in the heat flux in the piston of CCE over that of CE with its radius was shown in the Figure-8.

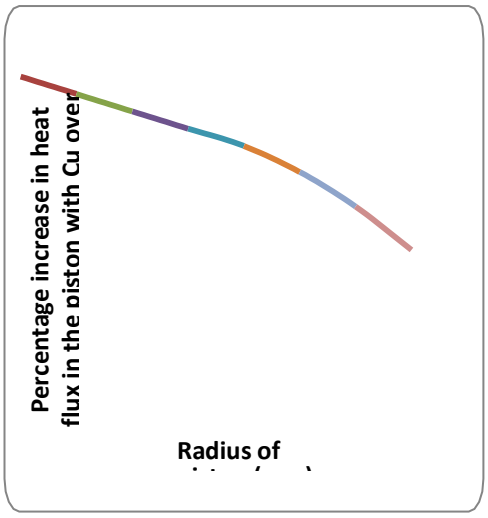


Figure 8 Variation of the % increase in the heat flux with radius of piston in CCE over CE.

As the radius of the piston increase, a marginal decrease in the heat flux(from19.6 % to 17.6 %) was noticed in the Figure-8. Since copper coating was done on the piston crown, maximum heat flux was concentrated at the centre of the piston crown. However, at the outer radius of the piston, heat flux was marginally lower, as the outer periphery of the piston is subjected to cooling by means of lubricating oil and fins.

The temperature at the inner side of the liner and cylinder head determined by FEM analysis were compared with the results obtained by experimentation, in order to ascertain the deviation of FEM results from experimental results.

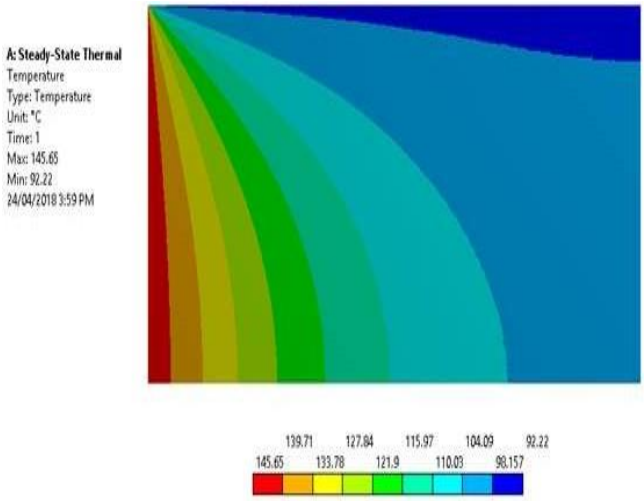


Figure-9,10 shows the isotherms of the lubricating oil between the piston and inner surface of liner of CE and CCE respectively

Figure 9 Isotherms from the finite element analysis in the lubricating oil between piston and inner surface of liner for CE

Figure 10 Isotherms from the finite element analysis in the lubricating oil between piston and inner surface of liner for CCE

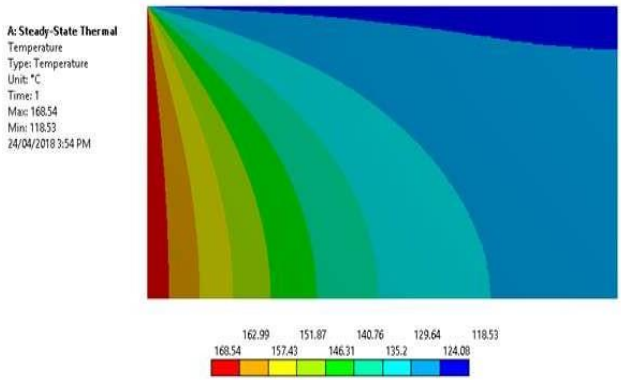
From the Figures-9,10 the temperatures of the lubricating oil varied from 92^0C to 145^0C for CE, while it varied between 118^0C to 168^0C for CCE respectively and found that temperatures are at safe limits for which lube oil does not deteriorate. The temperature limit for which lube oil does not deteriorate is 180^0C [12].

4. CONCLUSIONS

- Temperature at the top surface of the copper coated piston was 225^0C which is higher than 181^0C for the conventional engine.
- Heat flux is increased by17-19% along the radius of piston for the copper coated engine compared to conventional engine.
- Copper coating on the piston will not deteriorate the lubricating oil temperature as the lube oil temperature was in the safe limits between 118^0C to 168^0C,while it varied between 92^0C to 145^0C for conventional engine.

ACKNOWLEDGEMENTS

Authors thank authorities of Chaitanya Bharathi Institute of Technology, Hyderabad for facilities provided.



REFERENCES

[1] Khopkar, S.M., 2010, "Environmental Pollution Analysis," New Age International (P) Limited, Publishers, New Delhi, pp.180-190.

[2] Murali Krishna, M.V.S., and Kishor,K., 2008, "Performance Of Copper Coated Spark Ignition Engine With Methanol–Blended Gasoline With Catalytic Converter, "Indian Journal (CSIR) of Scientific and Industrial Research, 67,July,543–548.

[3] Kishor, K., Murali Krishna, M.V.S., Gupta, A.V.S.S.K.S., Narasimha Kumar, S., and Reddy, D.N, 2010, "Emissions From Copper Coated Spark Ignition Engine With Methanol Blended Gasoline With Catalytic Converter," Indian Journal of Environmental Protection,

30(3), pp.177–183

[4] Murali Krishna, M.V.S., Kishor, K., Murthy, P.V.K., Gupta, A.V.S.S.K.S., and Narasimha Kumar, S, 2010, ".Performance Evaluation Of Copper Coated Four Stroke Spark Ignition Engine With Gasohol With Catalytic Converter," International Journal of Engineering Studies, 2(4),pp.465–473.

[5] Murthy, P.V.K., Narasimha Kumar, S., Kishor,K., and Murali Krishna,M.V.S., 2011, "Aldehyde Emissions From Two-Stroke And Four-Stroke Spark Ignition Engines With Catalytic Converter Running On Gasohol," International Journal of Fuels and Lubricants, 50(2), pp.137-156.

[6] Murali Krishna,M.V.S., Kishor, K., Murthy, P.V.K., Gupta, A. V S. S. K .S. ,and Narasimha Kumar, S.,2011, "Comparative Studies On Performance Evaluation Of Four Stroke Copper Coated Spark Ignition Engine With Catalytic Converter With Alcohols,". International Journal of Advances in Engineering Research, 2(6), pp.*1–11*.

[7] V.R.Rajagopalan., S.Anand., D.Nagendra Kumar., V.Karunaharan., V.Lakshmi Narsimham.,"Investigations of In-Cylinder Heat Flux in a Single Cylinder, 4 Stroke, Air cooled, Spark Ignition Engine for Motor Cycle Application", SAE paper 320804,2015.

[8] Thet T.Mon, Rizalman Mamat, Nazri kamsah, "Thermal Analysis of SI-Engine using Simplified Finite Element Method", Proceedings of the world congress on engineering, Volume III,WCE- 2011.

[9] Murali Krishna,M.V.S., Kishor, K., Murthy, P.V.K., Gupta, A. V S. S. K .S. , and Narasimha Kumar, S.,2011, "Comparative Studies On Performance Evaluation Of Four Stroke Copper Coated Spark Ignition Engine With Catalytic Converter With Alcohols,". International Journal of Advances in Engineering Research, 2(6), pp.*1–11*

[10] Muhammet Cerit,Hakan S.Soyhan.,"Thermal analysis of a combustion chamber surrounded by deposits in an HCCI engine",Applied Thermal Engineering,volume-50,81-88,2013

[11] Kishor, k., "Investigations on Catalytically Activated Two Stroke Spark Ignition engine with alcohol blended gasoline", Ph.D. Thesis,J.N.T.university,Hyderabad,2014

[12] Heywood,J.B., "Internal Combustion Engine fundamentals", Thermo-chemistry of fuel air mixtures, Properties of working fluids ,Mc Graw-Hill Book Company, New York,1988.

ABOUT MRGI

Malla Reddy group of Institutions is one of the biggest conglomerates of hi-tech professional educational institutions in the state of Telangana, established in 2001 sprawling over 200 acres of land. The group is dedicated to impart quality professional education like pharmacy, Engineering & Technology, MCA, MBA courses. Our sole objective is to turn out high caliber professionals from those students who join us.

ABOUT MRCE

Malla Reddy group of Engineering (Formerly CM Engineering College) has been established under the aegis of the Malla Reddy Group of Institutions in the year 2005, a majestic empire, founded by chairman Sri Ch.Malla Reddy Garu. He has been in the field of education for the last 23 years with the intention of spearheading quality education among children from the school level itself. Malla Reddy College of Engineering has been laid upon a very strong foundation and has ever since been excelling in every aspect. The bricks of this able institute are certainly the adept management, the experienced faculty, the selfless non-teaching staff and of course the students.

ABOUT ICTIMES

ICTIMES started long back with its banner to promote the vision of future technologies that change the trends of life on this planet earth. Under this banner, the Department of Mechanical Engineering at MRCE organizes the ICFTME – International Conference on Future Technologies in Mechanical Engineering to provide a scholarly platform to ignite the spirit of Research and bring out the latent potential in teaching fraternity and student community. ICFTME accommodates major areas like, Applied Mechanics, Tribology, Advanced Manufacturing Process, Fluid Dynamics etc.,

ABOUT ICFTME

International Conference on Future Technologies in Mechanical Engineering **(ICFTME-2019)** will bring together innovative academicians, researchers and industrial experts in the field of Mechanical Engineering to a common forum. The idea of the conference is, for the scientists, scholars, engineers and students from the Universities across the world and the industry as well, to present ongoing research activities, and hence of foster research relations between the Universities and the industry with the rapid development of trends and studies in the fields concerned. **ICFTME-2019** will provide a heartwarming platform to researchers, scholars, faculty and students to exchange their novel ideas face to face together.

Estd :2005

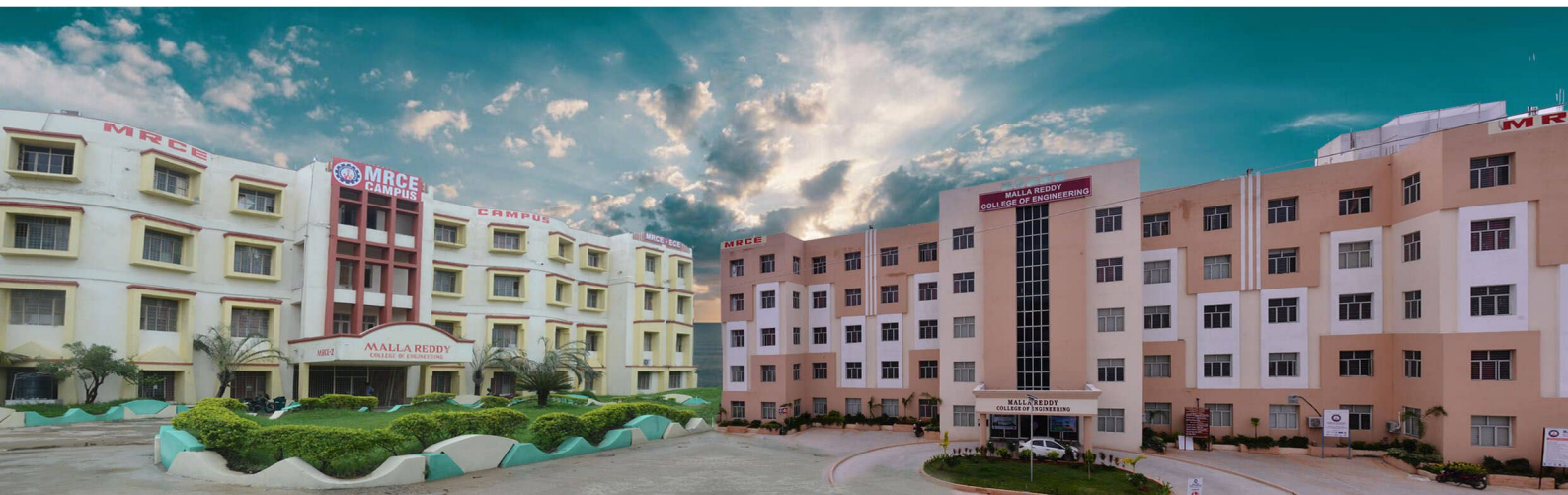
MALLA REDDY COLLEGE OF ENGINEERING

Approved by AICTE - New Delhi, Affiliated to JNTU - Hyderabad,Accredited by NBA & Accredited by NAAC.
ISO 9001:2015 Certified Institution, Recognition of College under Section 2(f) & 12 (B) of the UGC Act,1956.

Address: Maisammaguda, Dhulapally,(Post Via Kompally),Secunderabad - 500 100.

Ph: 040-64632248, 9348161222 ,9346162620. Email: principal@mrce.in

Website : www.mrce.in



ICTIMES 2019

5th INTERNATIONAL CONFERENCE ON

Trends in Information, Management, Engineering and Sciences

Proceedings of International Conference on Advances in Humanities Sciences and Management (ICAHSM)
ISBN No: 978-93-88808-63-7

Convenor
Dr. T. V. Reddy
Vice Principal & HOD-Dept. of H&S

Editor
Dr. Muthu Kumar
Professor - Dept. of MBA

Editor in Chief
Dr. P. John Paul
Principal

Organized by

Estd :2005

MALLA REDDY COLLEGE OF ENGINEERING

Approved by AICTE - New Delhi, Affiliated to JNTUH, Accredited by NBA & Accredited by NAAC.
ISO 9001:2015 Certified Institution, Recognition of College under Section 2(f) & 12 (B) of the UGC Act,1956.

website: www.mrce.in

ICTIMES - INTERNATIONAL CONFERENCE ON ADVANCES IN HUMANITIES SCIENCES AND MANAGEMENT (ICAHSM)

ZINCOXIDE PHOTO ELECTRO CHEMICAL CELL IN THE PRESENCE OF N-METHYL FORMAMIDE

Dr.T.V.Reddy, K.Saritha, S.Sirisha*
Department of Humanities & Sciences, Malla Reddy College of Engineering, Secunderabad, Telangana, India, drtvrchemistry@gmail.com

ABSTRACT

ZINCOXIDE(ZnO) is an n-type semiconductor which may be used as photo sensitive electrode and photo catalyst in the presence of organic substrates. ZnO photoelectrochemical (PEC) cell is constructed in the presence of N-methyl formamide (NMF). In oxygen atmosphere, ZnO electrode produces cathodic photocurrent which is attributed to the reduction of O_2 and anodic photocurrent is due to oxidation of NMF in the presence of N_2. A suitable mechanism is proposed for the electrode reactions.

Keywords: PEC CELL, ZnO, NMF, PHOTOVOLTAGE, PHOTOCURRENT

INTRODUCTION

The annual flux of solar energy falling on the earth is about 1.2 x 10^{21} Kcal.[1]. A large part of this (33%) is lost due to reflection by the atmosphere and the earth's surface. The rest of the energy is absorbed by the atmosphere (17%), oceans and seas (33%) and land (15%). Of the energy absorbed, only about 6 x 10^{7} Kcal./year (7 x 10^{-4}%) is converted into chemical energy via photosynthesis in plants and by micro organisms. However, only a small portion of this converted energy is utilized by mankind.

For the efficient utilization of solar energy that is available it is necessary to develop suitable methods for its storage. The photochemical conversion of solar energy into either electricity or chemical fuel is one of the most attractive pathways. The first observation of photoeffect at a metal or semiconductor electrolyte junction was reported by Becqurel [2]. A practical use for energy storage in the form of production of hydrogen from water was realized by Fujishima and Honda [3]. Since then extensive investigations have been carried out all over the world in the area of photochemical synthesis of energy rich compounds and photochemical decomposition of water.

Systems without semiconductors are found to be less efficient due to poor separation of products and reactants and occurrence of thermal back reactions between the products [4]. In the presence of semiconductors, the absorption of reactants on semiconductors and participation of electronic states of the semiconductor makes the system more efficient.

Usually the reactions in particulate systems and in PEC cells are compared on the basis of the products formed. In most of the cases the products are the same for both the systems and each particle of the semiconductor in the particulate system is treated as a micro PEC cell.

The study of PEC cells gained momentum with the first report of Fujishima and Honda [3] relating to the photo-electrochemical decomposition of water at an n-Tio_2 electrode illuminated by ultraviolet light. In 1975 Gerischer[5] showed the use of PEC cells for conversion of solar energy to electricity using reversible redox systems as the electrolytes.

An understanding of the process occurring inside the semiconductor on absorption of light and the charge transfer reactions that take place at the surface of the semiconductor in contact with the electrolyte is necessary for devising PEC cells for energy conversion.
Using semi conductor materials like ZnO, TiO_2, ZnO in the construction of PEC Cells received considerable attention. Any semi conductor to be effective in a PEC cell must be chemically inert and make use of solar energy in the visible region. **ZnO** is an n-type semi conductor with a band gap of 2.7 eV. This corresponds to utilization of 11.8% solar energy. **ZnO** is reported (1) as a stable semi conductor for electrolysis of water in the presence

of light. Photo catalytic activity of Z**n**O is also reported (2).

In the present study **ZnO** is used as a photo electrode in a PEC cell consisting of NMF Aqueous solution as electrolyte. Photo voltages and photo currents are measured. **ZnO** in the powder state is used as photo catalyst and the products are identified.

EXPERIMENTAL

PEC cell is constructed using polycrystalline **ZnO** paste. A thin layer of **ZnO** is deposited on a platinum foil and used as photo electrode. A platinum foil is used as counter electrode. A 1500W halogen lamp is used as light source. Electrolyte solution is a mixture of 0.1M KCl and 0.1M NMF.

ZnO,Pt/0.1 M NMF//0.1 M KCl/Pt
0.1 M KCl

here diagram to be inserted

RESULTS AND DISCUSSION

The sign of the open circuit voltage is found to depend upon gaseous atmosphere present at the electrode. In the absence of light the voltages are more positive in N_2 than in O_2. PEC cell produces a negative photo voltage and cathodic current in the presence of visible light and O_2. In the presence of N_2 a positive voltage is developed and anodic current is noted (table.1). If N_2 atmosphere is not maintained properly, the adsorbed O_2 on **ZnO** electrode surface initially produces cathodic current and then anodic current as shown in (Fig.1 & 2).

Table 1-Effect Of Gaseoius Atmosphere On Photo Voltage And Photo Current
Electrolyte: 0.1 M NMF in 0.1 M KCl (40ml)
Area of the electrode: 3 cm^2
Light source: 1000 watt halogen lamp

Gaseous atmosphere at		Open circuit photo-Voltage (mV)	Short circuit photo current (µA)	Nature of the photo-current
ZNO Electrode	Counter Electrode			
O_2	O_2	-150	3.1	Cathodic
O_2	N_2	-140	2.9	Cathodic
N_2	O_2	+45	0.5	Anodic
N_2	N_2	+43	1.0	Anodic

Photo voltage in O_2 reaches a limiting value within 15 minutes. It takes longer times, 150 minutes in N_2. When the light is switched off, the photo voltage of the cell in O_2 decays fast while in N_2 the decay is slow (Fig. 1).

The cathodic photocurrent in O_2 increases steeply on illumination. It reaches a maximum value, then decreases fast and attain a constant value (Fig.2) when the light is switched off, rapid decay of photo current is observed in both the atmospheres.

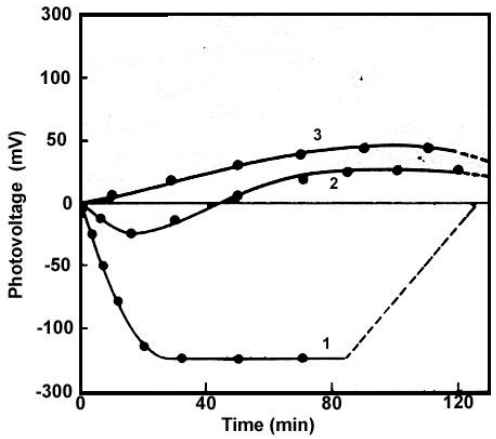


Fig.1-Variation of Photovoltage with time
1. Oxygen Atmosphere
2. Nitrogen Atmosphere (Bubbled for 60 minutes)
3. Nitrogen Atmosphere (Bubbled for 120 Minutes)
4. …….Light off

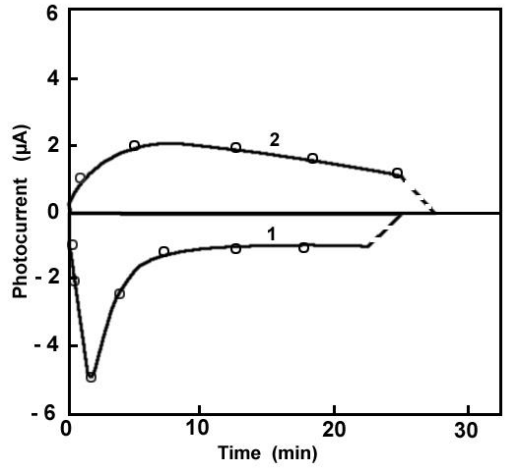


Fig2- Variation of Photocurrent with time
1. Oxygen Atmosphere
2. Nitrogen Atmosphere
3. …….Light off

Effect of intensity of light on photo voltage in O_2 atmosphere and 0.1M KCl and 0.1M N-Methyl formamide is studied. The photo voltage increases with light intensity and reaches a limiting value. Photo voltage is maximum around 360 nm, which corresponds to the band gap of Z**n**O (3.2 eV).

The plot of photovoltage verses log of light intensity is linear (Fig 3) which shows that **ZnO** behaves as a semiconductor (3).The generation of anodic photo current is due to oxidation reaction. In the present case cathodic photocurrent is observed only in the presence of oxygen. When O_2 is replaced by N_2 gas only anodic photo current is generated. Hence, cathodic photo current is due to reduction of O_2 to O_2^{--} Many Authors (8-10) observed similar effects on semi conductors. The super oxide Ion O_2^- is a good nucleophile and can react with carboxyl compounds (9). The O_2 reacts with N-methyl formamide to give CO_2 and $CH_3.NH_2$. CO_2 is identified at the electrode. However, CO_2 is not identified in the presence of N_2.

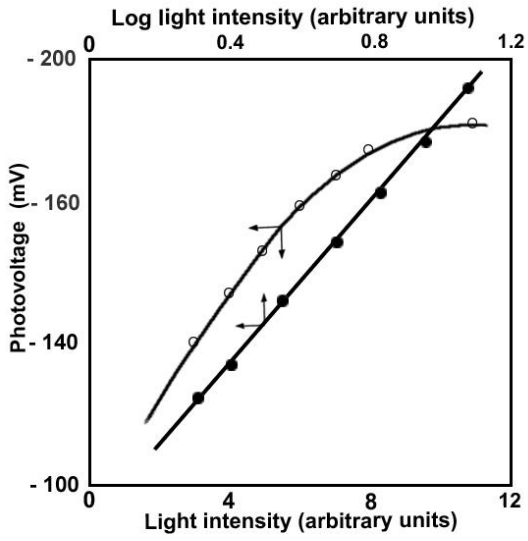


Fig3-Effect of intensity of light on photovoltage in oxygen atmosphere

Based on the experimental observations and the product analysis, a mechanism is proposed as given below. In reaction (2) the excited electrons in the conduction band flow to the surface of the electrode to react with adsorbed O_2 to give O_2^- and is responsible for the generation of cathodic photocurrent. The initial steep increase in photocurrent (Fig.3) is due to the availability of surface oxygen to react with the electrons as soon as illumination is started. The subsequent decrease may be due to poor electron hole separation and the oxidation of O_2^{--} radical by valence band holes (reaction 3). Detection of H_2O_2 provides additional evidence for the formation of O_2^{--} (12).

The other products, CO_2 and $CH_3.NH_2$ are formed only in the presence of oxygen and under illumination. Hence the O_2^{--} is expected to react with DMF to give CO_2 and $CH_3.NH_2$ as shown. Protonation of NMF (reaction 6) is well known in acidic medium and this facilitates O_2^{--} attack on the carbonyl carbon of the NMF.

The anodic nature of photocurrent in nitrogen

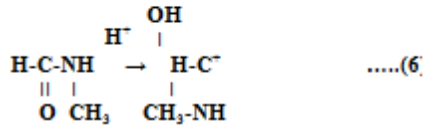
$$ZnO \xrightarrow{h\nu} ZnO^{(+)} + e^- \quad \text{....(1)}$$

$$e^- + O_{2\,(ads)} \rightarrow O_2^-{}_{(ads)} \quad \text{... (2)}$$

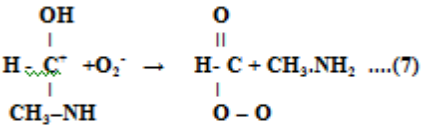
$$O_2^-{}_{(ads)} + ZnO^{(+)} \rightarrow O_2 + ZnO \quad \text{...(3)}$$

$$O_2^-{}_{(ads)} + H^+_{(ads)} \rightarrow HO_2^{\cdot} \quad \text{...(4)}$$

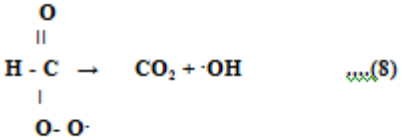
$$2HO_2^{\cdot}{}_{(ads)} \rightarrow H_2O_2 + O_2 \quad \text{...(5)}$$



.....(6)



....(7)



....(8)

$$2 \cdot OH \rightarrow H_2O_2 \quad \text{....(9)}$$

$$ZnO^- + OH^- \rightarrow ZnO + \cdot OH \quad \text{..... (10)}$$

When the electrode is in nitrogen atmosphere, the electrons that are excited to the conduction band flow into the bulk of the semiconductor to produce anodic photocurrent. The photocurrent increases initially, attaining a stable value and then decreases slowly. Initially there is a greater amount of band bending which efficiently separates photoproduced holes and electrons. As the illumination progresses, the separation becomes less efficient due to decrease in band bending, resulting in recombination of electrons and holes. Hence the anodic photocurrent after reaching a maximum value starts decreasing.

It has not been possible to completely eliminate either oxidation or reduction at the electrode in a particular gaseous atmosphere. Both reactions can take place simultaneously and whichever dominates decides the sign of the photovoltage. In oxygen, reduction takes place by conduction band electrons along with the oxidation of NMF by valence band holes during illumination (14). But the rate of reaction with holes is very slow while the reduction of oxygen is faster. Hence negative photovoltage dominates in this case. In nitrogen the presence of minute amounts of oxygen, like surface oxygen leads to an initial generation of negative photovoltage which gradually becomes positive. Only prolonged purging of the electrolyte with nitrogen completely eliminates negative photovoltage (Fig.2).

CONCLUSION

Development of photoelectrochemical cells less sensitive to air and not requiring hermetic seals will decrease the fabrication costs as well as the technological problems. One can consider the Z**n**O based photoelectrochemical cell which is not sensitive to air. But the efficiency of such a cell is very low. Investigations are in progress to improve the cell efficiency and electrode stability.

References

1. K.I.Zamarev and V.N.Parmon, Catl Rev.Sci.Engg, 22, (1980) 261.
2. E.Becqurel C.R. , Acad.Sci, 9(1839) 561
3. Dr.Fujishima and K.Honda, Nature, 238 (1972) 37
4. H.Gerischer, J.Electro and Chem. 52, (1980) 2 649
5. H.Giresher, J.Electro and Chem 58 (1975) 263
6. Buttler M A, Nasby RD & Quinn R K, *Solid Stat Commun*, 19 (1976) 1011.
7. Ravindranathan Thampi K, Varahala Reddy T, Ramakrishan V & Kuriacose J C, *Abstracts of X International Conference on Photochemistry*, Greece, published *in J Photochem*, 17 (1-2) (1981) 183.
8. Davidson R S, Meek R R & Slater R M, *J Chem Soc Faraday I*, 75 (1979) 2507.
9. Miyake M, Yoneyama H & Tamura H, *J Catal*, 58 (1079) 22.
10. Sprunken H R, Schumacher R & Schimdler R.N, *Ber Bunsenges phys Chem*, 84 (1980)1040.
11. Dunn W W, Aikawa Y & Bard A J, *J Electrochem Soc*, 128 (1981) 222
12. Hardee L K & Bard A J, J *electrochem Soc*, 122 (1975) 739.
13. Harbour J R & Hair M L, *J Phys Chem*, 83 (1979) 652.
14. Magnow F, Botempalli G & Andreazzisedean M M, *J Electroanalchem*, 97 (1979) 85.

Time Dependent Assignment problem and its extensions for construction project scheduling

1) **Animoni Nagarju**,
Associate Professor , Department of Mathematics , Malla Reddy Institute of Technology and Sciences,Hyderabad, Telangana, India
animoni_nagaraju@yahoo.co.in

2) **S Srilakshmi Alla**
Assistant Professor , Department of Mathematics , Malla Reddy Institute of Technology and Sciences,Hyderabad, Telangana,India.
srilakshmialla1984@gmail.com

3) **Shashi kumar Jakkaraju**
Assistant Professor , Department of Engineering physics , Malla Reddy College of Engineering Hyderabad,Telangana,India.
jakkarajushashi@gmail.com

ABSTRACT:

The time dependent assignment problem consists of allocating renewable resources (construction equipment, crews, or contractors, machines , jobs and persons) of limited availability to a set of activities, this perform a subtask is limited to specific intervals in time. The classical model for this problem objective is to minimize the total time or cost of completing all activities with the assumption that each activity is assigned to one particular resource at perticular period. A time-dependent task is a task requiring multiple renewable resources to perform separate subtasks simultaneously or within some predetermined margin where resource availability to perform a subtask. This paper systematizes and describes extensions of these assumptions, considering the effects of task sequence: parallel, serial and hybrid (modeled by means of network methods). This study proposes algorithms for the solution of presented models, which can be used in construction project scheduling.

KEYWORDS: assignment problem, project scheduling, mathematical modeling, renewable resources, bottleneck assignment problem

I. INTRODUCTION

The time dependent Assignment problems (selection and allocation of resources to the jobs or machines at perticular time period) are one of the primary task in construction process synchronization at particualr times. In its classical formulation *n* teams (i.e crews, sets of construction plant, contractors, and samples) are assigned to *n* activities at particualr r times, in order to minimize the total time or cost of completing all of them. Arrangement of these activities depends on technological and organizational requirements. These requirements concurrently affect the allocation constraints, such as assigning only one job to each machine at particualr periods. With respect to technological constraints, options for conducting construction processes in parallel, in series, or in a hybrid way may exist. Parallel processing consists of the simultaneous execution of activities by different resources in separate building units (work zones) at particualr times. The greatest advantage of this technique is the shortest project time span *T* – when compared with other approaches. The drawbacks of the parallel technique are: the lack of teams' work continuity and unleveled daily demand for building materials or plant as per duration.

Serial processing consists of performing a sequence of processes in one work zone in the seasonally. The advantage of this method is the lowest maximum level of daily employment of renewable resources and daily usage for building materials. Each activity may be realized by a different crew, but the total duration is incommensurately long. Another disadvantage,

similar to the parallel technique is discontinuity of the team's work and unleveled daily consumption of resources. Hybrid processing is a combination the two previous approaches seasonally. The precedence constraints, modeled by means of network techniques, enable serial processing of some activities (on the same network path) and, simultaneously, concurrent processing of other jobs on parallel paths. The two optimization approaches are used for resource management in project networks: allocation of limited resources (in order to minimize the project makespan) and leveling resource requirements profile (in order to improve economicefficiency). This study considers different processing options for assignment problem

formulation. It has been assumed that an activity requires one resource type for its execution (crew, contractor), and the resource can perform one activity at a time with particular time periods seasonally. In the case of parallel processing, each activity has to be carried out by a different crew. With serial processing, each activity can, but does not have to be conducted by a different crew. Thus, in the case of hybrid processing,

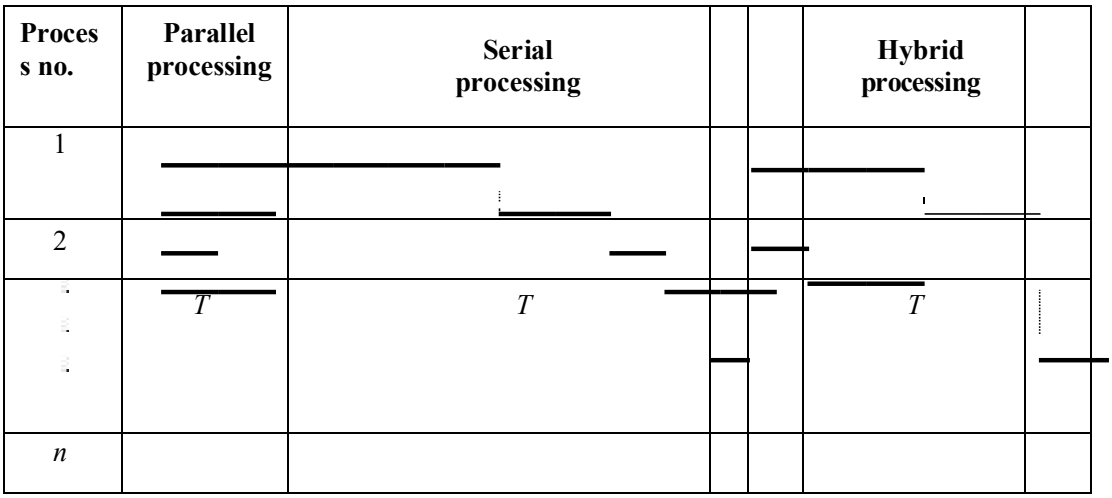


Fig. 1. Comparisons of different processing options in processes scheduling

parallel activities have to be entrusted to different crews, and activities scheduled in sequence can but do not have to be carried out by different crews. A discrete time/cost/resource function implies the representation of an activity in different modes of operation. For each activity *i* (construction process), a set of modes of execution is defined. Each mode, is described by the following parameters: $t_{ij} \in N$ – duration of the activity *i* realized by crew *j* and $k_{ij} \in R$ – cost of the activity *i* realized by crew+ *Jat particular time period k*. The problem consists of choosing the optimal process modes in order to minimize the project duration or cost. Binary linear programs are developed to model the assignment problem for different processing options and recommended approach solutions are presented.

II FORMULATION

1 Assignment problem formulation for parallel processing

The classical assignment problem was formulated in 1952 by D.F. Votaw and A. Orden
[1] as a type of transportation problem. Present extended study in on time dependent assignment problem. The mathematical model of the time dependent assignment problem for parallel processing can be described by an objective function minimizing the total cost of realizing processes at particualr times
:

$$\text{MinZ=Min}\left(\sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{k=1}^{s} k_{ijk}x_{ijk}\right)$$

and the following constraints:

$$\sum_{i=1}^{n} x_{ijk} = 1, \quad j = 1, 2, ..., n, \text{ k=1,2…r} \qquad (2)$$

$$\sum_{j=1}^{n} x_{ijk} = 1, i = 1, 2, ..., n, \text{ k=1,2…r} \qquad (3)$$

$x_{ijk} \in \{0,1\}, i = 1, 2, ..., n, j = 1, 2, ..., n$, k=1,2…r
where:

x_{ijk} – a binary variable for modeling a decision of selecting processes' modes; the variable assumes the value of 1 if the activity *i* is to be executed by the resource (crew) *j*, and equals 0 in the other case. According to equations (2) and (3), considering equal numbers of crews and activities, each crew may be assigned only to one process and each process may be realized by only one crew at the particular time period.

H.W. Kuhn [2] in 1955 created the Hungarian method – exact algorithm for solving the model. He combined the ideas of two Hungarian mathematicians: D. König [3] and J. Egerváry [4]. The method finds an optimal assignment for a given square cost matrix, and consist of five steps [2, 5–7]: with the considered time dependent periods.

Step 1 – Subtract the smallest entry in each row from all the entries of its row.

Step 2 – Subtract the smallest entry in each column from all the entries of its column.

Step 3 – Construct a minimal number of lines, which covers all the zero entries of the cost matrix with k_{ij}.

Step 4 – If number of covering lines is *n*, then it is complete. Otherwise, proceed to Step 5.

Step 5 – Determine the smallest, uncovered entry and subtract it from uncovered rows, and then add this entry to each covered column. Repeat Step 3.

The bottleneck assignment problem is one of the extensions of the classical formulation. It consist of the minimization of the project makespan – maximal duration of processes realized in parallel (each demanding a different crew):

$$\text{Min z: } z = \max \sum_{j=1}^{m} \sum_{i=1}^{n} \sum_{k=1}^{s} k_{ijk} x_{ijk} \quad (5)$$

subject to the same constraints and definitions as in the classic time dependent assignment problem. In 1959 O. Gross [8] created an algorithm which is used for solving this kind problem. It may be described as follows: time dependent assignments

Step 1 – Begin with any of feasible solution, e.g. $x_{ii} = 1$, I = 1, 2, …, n, set of chosen entries – $B = \{t_{ij} : x_{ij} = 1\}$.

Step 2 – Compute $V = \max\{t_{ij} : x_{ij} = 1\}$.

Step 3 – Locate a cycle, which begins and ends at V as follows:

Step 3a – from V go to entry in its column with cost less than V,

Step 4 – Reverse the assignments along entries, which take part in previous steps. Let $x'_{ij} = 1 - x_{ij}$ for participating entries in Step 2. Proceed to Step 2.

2.Assignment problem formulation for serial processing

After completing each serial process, the renewable resources become available again and may be assigned to succeeding activities with the time periods. Therefore, the constraint (2) in the classical formulation doesn't hold in

this case. For each process, the whole set of crews is considered as the feasible solution space – in the problem of minimizing project duration as well as cost, aappropriate modes should be selected with the shortest duration for each process or with minimal cost, respectively.

Assuming that the processes sequence will be repeated continuously in d identical work zones (units), the objective function of minimizing project duration should be modified as follows:

$$\text{Min } z : z = (d-1)\max\{t_{ijk} \cdot x_{ijk}\} + \sum_{j=1}^{m} \sum_{i=1}^{n} \sum_{k=1}^{s} t_{ijk} x_{ijk}$$

The mathematical model (converted to linear form) with binary variables for many practical instances may be solved using any commercial optimization software.

3.Assignment problem formulation for hybrid processing

A construction project can be modeled as an activity–on–node network. Precedence relations between activities are modeled by a graph $G = <V, E>$, directed and acyclic, with

a single initial node and a single final node, where $V = \{1, 2, \ldots, n\}$ is a set of activities with dependent time periods, the edges (or arcs) $E \subset V \times V$ represent precedence relations between activities. R is the set of resources – crews or contractors – available to the project. Variables s_i, $\forall i \in V$, stand for activities' start times

. Resources can be assigned to a number of processes, but not at the same time. Therefore, a set of processes' pairs $J \; V \times V$ can be defined, which can potentially be executed in parallel ($(u, v) \in J \Leftrightarrow u < v$ and activities u and v do not lie on the same path of the project network).

In the case that the resource j ($x_{ij} = 1 \wedge x_{vj} = 1$) is assigned to a pair of processes $(u, v) \in J$, these

Processes cannot run at the same time,

but have to be completed in sequence. The sequence is modeled by means of binary variables: $y_{uv} \in \{0, 1)$, defined for $\forall (u, v) \in J$. The variable y_{uv} equals 1 if the activity u is to be completed before activity v, and it equals 0 in the other case.

The decision making process is aimed at selecting options of resource assignment and scheduling them in such a way that project duration is minimal. To solve the problem, a mixed integer (binary) linear program is developed to model the construction project scheduling problem. The mathematical model used for this problem is described as follows:

$$\min T: T = s_n + D_n \qquad (7)$$

$$D_i = \sum_{j\in R}\sum_{k\in R} t_{ijk}\cdot x_{ijk},\ \forall i\in V \qquad (8)$$

$$\sum_{j\in R}\sum_{k\in R} x_{ijk} = 1,\ \forall i\in V \qquad (9)$$

$$s_1 = 0 \qquad (10)$$

$$s_i + D_i \le s_l,\ \forall (i,l)\in E \qquad (11)$$

$$s_u + D_u \le s_v + M\cdot(1-y_{uv}) + M\left(2 - x_{uj} - x_{vj}\right),\ \forall (u,v)\in J, \forall j\in R \qquad (12)$$

$$s_v + D_v \le s_u + M\cdot y_{uv} + M\cdot\left(2 - x_{uj} - x_{vj}\right),\ \forall (u,v)\in J, \forall j\in R \qquad (13)$$

$$s_i \ge 0,\ \forall i\in V \qquad (14)$$

$$x_{ijk}\in\{0,1\}, \forall i\in V, \forall j\in R\ \forall k\in R \qquad (15)$$

$$y_{uv}\in\{0,1\}, \forall (u,v)\in J \qquad (16)$$

The objective function (7) minimizes total project duration. Equation (8) determines duration D_i of a activity i – it has been introduced as an auxiliary formula to simplify the formulas (7) and (11)–(13). According to condition (9), each activity can be executed in only one way – as selected from available options. Execution of the first activity of the project (i.e. a activity that has no predecessors) starts at the moment of 0 (10). Condition (11) defines the successors' start dates as "not earlier than their predecessors have been completed with time dependent condition". Formulas (12) and (13) are introduced to define process start times $(u, v) \in J$. If these processes are not to be executed by the same resource j, ($x_{ij} = 0 \vee x_{vj} = 0$), both of these conditions are automatically met (M is an arbitrarily assumed, sufficiently large constant), and the processes may run concurrently. If the same resource j is assigned to them ($x_{uj} = 1 \wedge x_{vj} = 1$), and if the variable y_{uv} assumes the value of 1, then (in accordance with condition (12)), process v can only start after process u has been completed; in this case, condition (13) is automatically fulfilled. If the variable y_{uv} equals 0, then process v must be completed before, u has been started – according to condition (13) and condition (12) is met automaticall

The problem of project cost minimization is trivial and in the optimal solution the crews with the lowest cost for each process are assigned to realize it. Because the project duration for this assignment may be unacceptably long, the objective could be modified as follows

$$\text{Min k: } k = \sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{k=1}^{s} k_{ijk}x_{ijk} \qquad (17)$$

and the following constraint needs to be added to assure not exceeding the deadline T_d

$$s_n + D_n \le T_d \qquad (18)$$

III. EXAMPLE

Durations (in days) and costs (in monetary units) for completing a particular processes by particular crews in an example project is presented in matrices **T** and **K** respectively.

$$\mathbf{T} = \begin{bmatrix} 9 & 7 & 6 & 5 & 4 \\ 6 & 5 & 8 & 6 & 4 \\ 3 & 5 & 2 & 5 & 5 \\ 4 & 4 & 4 & 3 & 4 \\ 6 & 5 & 8 & 7 & 6 \end{bmatrix},$$

$$\mathbf{X}_{\mathrm{I}}^{1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

The hybrid precedence relations of the processes is settled according to the technological constraints and shown in Fig. 2.

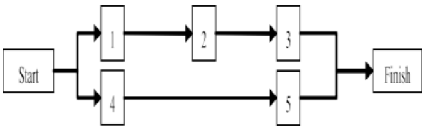


Fig. 2.

There are two optimal solutions of the assignment problem for parallel processes (with minimal project cost of 19 monetary units) obtained using the Hungarian algorithm:

$$\mathbf{X}_{\mathrm{I}}^{1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix},$$

and

$$\mathbf{X}_{\mathrm{II}}^{1} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

The Hungarian algorithm prompts that there are three different optimal solutions obtained for the original assignment problem with minimal sum of processes' durations (20 days):

$$\mathbf{X}^{2} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{X}^{2} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{X}^{2} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

The solution ^{1} is also obtained by means of the algorithm by Gross. The minimal.duration for the bottleneck assignment problem is 5 days. Because the solutions for both the bottleneck and the original assignment problems are the same, so this is also the optimal crew assignment for the sequence of processes repeated in different units (regardless of the units number).

The optimal solution for crew assignment processes with hybrid precedence constraints are as follows:

The minimal project duration for a given network is 10 days. Solution of the mathematical model (7)–(16) based on the above input values was found by means of LINGO 12.0 Optimization Modelling Software by Lingo Systems Inc.

IV. CONCLUSION

Scheduling with the allocation of constrained resources, particularly for skilled labor, is a major challenge for almost all construction projects with time dependent constraint. Most existing techniques for project scheduling consider a single–skilled resource strategy where each worker or crew is assumed to be of one particular trade. This strategy may lead to inefficiencies in labor utilization, which can also be reflected in increased project durations and unnecessary costs. In practice each construction contractor or crew can possess several skills at different proficiency levels,
i.e. they are able to perform more than one type of work (construction process), each at specified times and costs. The assumption that each worker may possess multiple skills which could allow them to participate in any activity that fits one of their skills, can improve project efficiency in terms of

project cost and duration with considered time durations.

The resources assignment models presented in the paper can help managers in determining a strategy for crew or bids selection.

For the case of parallel processing there are exact algorithms available for solving assignment problems with low computational effort. For hybrid processing options, the problems analyzed in this paper can be considered as an extension of the Resource– Constrained Project Scheduling Problem (RCPSP) which is NP–hard. Exact algorithms may be not efficient to solve complex practical problems with considered time dependent constraint , therefore developing heuristic solving procedures is recommended for further research.

REFERENCES

[1] Votaw D.F., Orden A., The personnel assignment problem, Symposium on Linear Inequalities and Programming, SCOOP 10, US Air Force, 1952, 155-163.

[2] Kuhn H.W., The Hungarian method for the assignment problem, Naval Research Logistics Quarterly 2 (1&2), 1955, 83-97.

[3] König D., Über Graphen and ihre Anweendung auf Determinantentheorie und Mengenlehre, Math. Ann. 77, 1916, 453-465.

[4] Egerváry J., Combinatrional Properties of Matrices, ONR Logistics Project, Priencton 1953.

[5] Jaworski K.M., Metodologia projektowania realizacji budowy, Wydawnictwo Naukowe PWN, Warszawa 2009, 159-162.

[6] Kuhn H. -W., A tale of three eras: The discovery and rediscovery of the Hungarian Method, European Juournal of Operational Research 219, 2012, 641-651

[7] www.math.harvard.edu/archive/20_spring_05/handouts/assignment_overheads.pdf

[8] Gross O., The Bottleneck Assignment Problem, P-1630, The Rand Corporation, Santa Monica, California 1959.

[9] Jaworski K.M., Podstawy Organizacji Budowy, Wydawnictwo Naukowe PWN, Warszawa 2011.

[10] Nagaraju.A., Variant constraint of time dependent time minimization assignment problem with minimize objective- A Lexi search approach, International e journals of mathematics and engineering 221.,2013,2170-2184.

[11] Syakinah_Faudzi, Syariza_Abd ul-Rahman, and Rosshairy Abd Rahman, An Assignment Problem and Its Application in Education Domain: A Review and Potential Path, Advances in Operations research, 2018, Article ID 8958393 ,https://doi.org/10.1155/2018/8958393.

SOME RESULTS ON CLOSED INTERVAL-VALUED M-POLAR FUZZY GRAPHS

Ramakrishna Mankena [1, *], T.V. Pradeep Kumar [2]

[1, *] Assistant professor, Department of Mathematics, Malla Reddy College of Engineering, Hyderabad, India, E-mail: [1]rams.prof@gmail.com

[2] Department of Mathematics, University College of Engineering, AcharyaNagarjuna University, Nagarjuna Nagar, India, E-mail: [2]pradeeptv5@gmail.com

Abstract:

In this paper, we introduced a new concept of closed interval-valued m-polar fuzzy (CIVmPF) graph and investigated some of its properties. Also defined degree and total degree of an edge and obtained equivalence condition for edge regular CIVmPF graph and totally edge regular CIVmPF graphs.

Keywords: CIVmPF graph, Edge degree, Total edge degree, Complete graph.

Section 1: Introduction

In non-deterministic real life situations free body diagrams are using as a set of nodes connected by lines representing pairs are good problem solving tools. In this connection graphs developed using discrete parameters to solve problems related to vague and uncertain real life situations are demonstrated by Zadeh [28] in 1965. Problems related to networks that demand intuitive data analysis technique are solved by interval valued fuzzy sets introduced by Zadeh [29]. The limitations of traditional model were overcome by the introduction of bipolar fuzzy concept in 1994 by Zhang [30, 31]. This was further improved by Chen et al. [6] to m-polar fuzzy set followed by Kaufmann [13] fuzzy graph. Rosenfeld [21], Bhutani et al. [4, 5], Samanta and Pal [22], Sunitha and Mathew [25] extended the fuzzy graph theory to accommodate various complex problems in planar graphs. Isomorphism and complement of bipolar fuzzy graph and interval-valued fuzzy graphs were prefaced by Talebi and Rashmanlou et al. [26, 27]. While Rashmanlou et al. [19, 20] studied the categorical properties of bipolar fuzzy graphs. Ghorai and Pal [7-12] introduced single m-values to m-polar fuzzy graphs followed by extensive work on planarity in vague graphs. Self centered interval valued fuzzy graphs and interval valued fuzzy line graphs were developed by Akram et al. [1-3]. Intutionistic fuzzy tolerance graph theory

models were introduced by Sahoo and Pal [23, 24] to understand tolerance, product and degree of intutionistic fuzzy graph. Genus value of m-polar fuzzy graphs was discussed by Mandal et al. [16]. Ramprasad et al. [17, 18] studied the properties of product m-polar fuzzy line graph, product m-polar fuzzy intersection graph and h-morphism between two m-polar fuzzy graphs. Mordeson and Nair [14, 15] studied fuzzy hypergraphs and some operations on it.

This paper attempts to develop theory to analyze parameters combining concepts from m-polar fuzzy graphs and interval valued fuzzy graphs as a unique effort. The resultant graph is turned CIVmPF graph and studied properties on it. CIVmPF graph theories have several applications in different fields such as image capturing, image segmentation, networking, data mining, planning (air lines connectivity, landscape connectivity etc.), scheduling and clustering.

Section 2: Preliminaries

An interval number, $\bar{N}$, is defined as an interval $\left[N_S, N_H\right]$, where $0 \le N_S \le N_H \le 1$. The set of all interval in between 0 and 1 is denoted by $T[0,1]$. The interval $[N,N]$ is identified with the numbers $N \in [0,1]$. For interval number $\bar{N}_j = \left[N_{S_j}, N_{H_j}\right] \in T[0,1]$ and $j = 1, 2, \cdots$.

We define

$\inf \bar{N} = \left[\wedge_{j\in Z^+} N_{S_j}, \wedge_{j\in Z^+} N_{H_j}\right]$ and

$\sup \bar{N} = \left[\vee_{j\in Z^+} N_{S_j}, \vee_{j\in Z^+} N_{H_j}\right]$, where Z^+ is the set of positive integers. We define

i) $\bar{N}_1 \le \bar{N}_2 \Leftrightarrow N_{S_1} \le N_{S_2}$ and $N_{H_1} \le N_{H_2}$.

ii) $\bar{N}_1 = \bar{N}_2 \Leftrightarrow N_{S_1} = N_{S_2}$ and $N_{H_1} = N_{H_2}$.

iii) $\bar{N}_1 < \bar{N}_2 \Leftrightarrow \bar{N}_1 \le \bar{N}_2$ and $\bar{N}_1 \ne \bar{N}_2$.

iv) $m\bar{N} = \left[mN_S, mN_H\right]$, where $0 \le m \le 1$.

Definition 2.1. A *closed interval-valued m-polar fuzzy set* Q in V is defined by

$$Q = \left\{ \begin{matrix} \left(q, \left[N_{S_1(q)}, N_{H_1(q)}\right]\right), \left(q, \left[N_{S_2(q)}, N_{H_2(q)}\right]\right), \\ \cdots, \left(q, \left[N_{S_m(q)}, N_{H_m(q)}\right]\right) \end{matrix} \right\},$$

where $N_{S_j(q)}$ and $N_{H_j(q)}$ are a lower and upper limit of the j^{th} fuzzy interval $j = 1, 2, \cdots m$.

Definition 2.2. A *closed interval-valued m-polar fuzzy relation* T on a set E is defined as follows: for $j = 1, 2, \cdots m$

$$NE_{S_j(\alpha, \beta)} \leq N_{S_j(\alpha)} \wedge N_{S_j(\beta)}$$

$$NE_{H_j(\alpha, \beta)} \leq N_{H_j(\alpha)} \wedge N_{H_j(\beta)}$$

Let $G = (V, E)$ be a graph where V (non-empty set) is called vertex set and E (empty or non-empty set) is called edge set. If no edge incident with a vertex, then the vertex is said to be isolated vertex, otherwise, it is said to be non-isolated vertex.

Definition 2.3. An *m-polar fuzzy graph* $G = (V, W, F)$ with underlying crisp graph $G^* = (V, E)$

Where $W : V \to [0, 1]^m$ is an m-polar fuzzy set in V and $F : V \times V \to [0, 1]^m$ is an m-polar fuzzy set in $V \times V$ such that for each $j = 1, 2, \cdots m$,

$p_j \circ F(\alpha, \beta) \leq \inf \{ p_j \circ W(\alpha), p_j \circ W(\beta) \}$ for all $(\alpha, \beta) \in V \times V$ and $F(\alpha, \beta) = 0$ for all $(\alpha, \beta) \in (V \times V - E)$, ($O = \langle 0, 0, \cdots, 0 \rangle$ is the least element in $[0, 1]^m$). $p_j \circ W(\alpha)$ and $p_j \circ F(\alpha, \beta)$ represents the j^{th} component of the degree of membership value of a vertex 'α' and '(α, β)' edge of m-polar fuzzy graph.

Section 3: Regularity on CIVmPF graph

In m-polar fuzzy graph, each vertices or edges have m components and those components are fixed. But, these components may be interval. Using this idea, CIVmPF graph has been introduced in this section.

Definition 3.1. A *CIVmPF graph* $G = (V, W, F)$ with underlying crisp graph $G^* = (V, E)$ where $W : V \to T[0, 1]^m$ is aCIVmPF set in V where $p_j \circ W(\alpha) = \left[N_{S_j(\alpha)}, N_{H_j(\alpha)} \right]$ for each $j = 1, 2, \cdots m$ and $F : V \times V \to T[0, 1]^m$ is a closed interval-valued m-polar fuzzy relation in $V \times V$, where $p_j \circ F(\alpha, \beta) = \left[NE_{S_j(\alpha,\beta)}, NE_{H_j(\alpha,\beta)} \right]$, for each $j = 1, 2, \cdots m$, and satisfying the following for all $(\alpha, \beta) \in V \times V$ and for each $j = 1, 2, \cdots m$,

$$NE_{S_j(\alpha, \beta)} \leq N_{S_j(\alpha)} \wedge N_{S_j(\beta)},$$

$$NE_{H_j(\alpha, \beta)} \leq N_{H_j(\alpha)} \wedge N_{H_j(\beta)},$$

and $F(\alpha, \beta) = (0, 0)$ for all $(\alpha, \beta) \in (V \times V - E)$. Here, $\left[N_{S_j(\alpha)}, N_{H_j(\alpha)} \right]$ and $\left[NE_{S_j(\alpha,\beta)}, N_{H_j(\alpha,\beta)} \right]$ represents j^{th} component

of degree of membership value of a vertex 'α' and (α, β) edge of a CIVmPF graph respectively.

Example 3.1. Here, an example of *interval-valued 3-polar fuzzy graph* has given below.

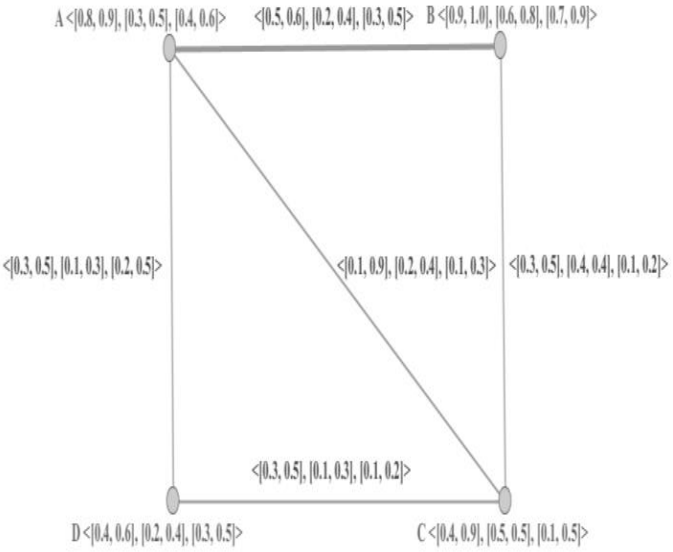


Figure 1: Interval-valued 3-polar fuzzy graph G

Defination 3.2. ACIVmPF graph
$G=(V, W, F)$ is said to be *complete* if, for each $j=1, 2, \cdots m$ and for each $\alpha, \beta \in V$

$$NE_{S_j(\alpha, \beta)} = N_{S_j(\alpha)} \wedge N_{S_j(\beta)},$$

$$NE_{H_j(\alpha, \beta)} = N_{H_j(\alpha)} \wedge N_{H_j(\beta)}.$$

Example 3.2. Here, an example of *complete closed interval-valued 3-polar fuzzy graph* has given below.

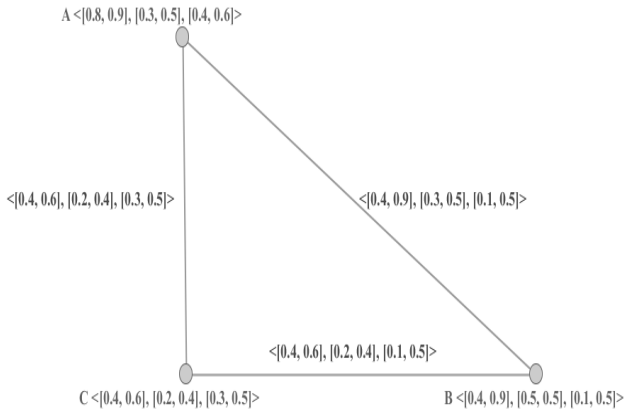


Figure 2: A complete Interval-valued 3-polar fuzzy graph G

Definition 3.3. ACIVmPF graph
$G=(V, W, F)$ with underlying crisp graph $G^*=(V, E)$ is said to be *strong* if, for each $j=1, 2, \cdots m$ and for each $(\alpha, \beta) \in E$

$$NE_{S_j(\alpha, \beta)} = N_{S_j(\alpha)} \wedge N_{S_j(\beta)},$$

$$NE_{H_j(\alpha, \beta)} = N_{H_j(\alpha)} \wedge N_{H_j(\beta)}.$$

Example 3.3. Here, an example of *strong interval-valued 3-polar fuzzy graph* has given below.

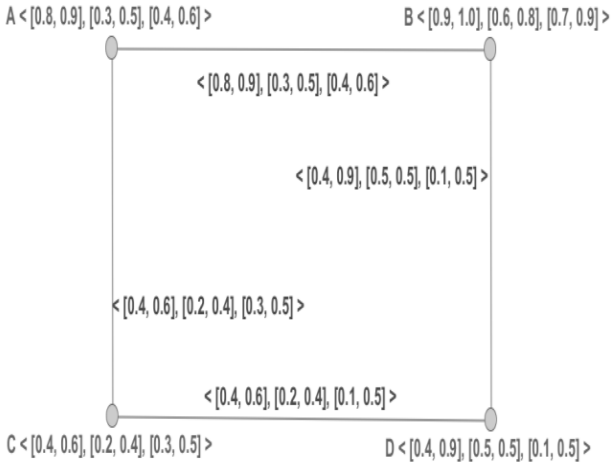


Figure 3: A strong Interval-valued 3-polar fuzzy graph G

Definition 3.4. The *degree of a vertex* α in aCIVmPFgraph is defined as

$$d_G(\alpha)=\left\langle \begin{array}{l} \sum_{(\alpha,\beta)\in E,\beta\in V}\left[NE_{S_1(\alpha,\beta)},NE_{H_1(\alpha,\beta)}\right], \\ \sum_{(\alpha,\beta)\in E,\beta\in V}\left[NE_{S_2(\alpha,\beta)},NE_{H_2(\alpha,\beta)}\right], \\ \cdots, \sum_{(\alpha,\beta)\in E,\beta\in V}\left[NE_{S_m(\alpha,\beta)},NE_{H_m(\alpha,\beta)}\right] \end{array} \right\rangle$$

.

Definition 3.5. The *degree of an edge* $(\alpha,\beta)\in E$ in aCIVmPF $G=(V,W,F)$ is defined as

$$d_G(\alpha,\beta)=d_G(\alpha)+d_G(\beta)-2\left\langle\left[NE_{S_1(\alpha,\beta)},NE_{H_1(\alpha,\beta)}\right],\left[NE_{S_2(\alpha,\beta)},NE_{H_2(\alpha,\beta)}\right],\cdots,\left[NE_{S_m(\alpha,\beta)},NE_{H_m(\alpha,\beta)}\right]\right\rangle$$

Definition 3.6. The total degree of an edge $(\alpha,\beta)\in E$ in aCIVmPFgraph $G=(V,W,F)$ is defined as

$$td_G(\alpha,\beta)=d_G(\alpha)+d_G(\beta)$$
$$-\left\langle \begin{array}{l} \left[NE_{S_1(\alpha,\beta)},NE_{H_1(\alpha,\beta)}\right],\left[NE_{S_2(\alpha,\beta)},NE_{H_2(\alpha,\beta)}\right], \\ \cdots,\left[NE_{S_m(\alpha,\beta)},NE_{H_m(\alpha,\beta)}\right] \end{array} \right\rangle$$

.

Throughout this paper, we considered

$<\left[\delta_1^-,\delta_1^+\right],\left[\delta_2^-,\delta_2^+\right],\cdots,\left[\delta_m^-,\delta_m^+\right]>$ as

$<\delta_1,\delta_2,\cdots,\delta_m>$,

$<\left[h_1^-,h_1^+\right],\left[h_2^-,h_2^+\right],\cdots,\left[h_m^-,h_m^+\right]>$ as

$<h_1,h_2,\cdots,h_m>$,

$<\left[\gamma_1^-,\gamma_1^+\right],\left[\gamma_2^-,\gamma_2^+\right],\cdots,\left[\gamma_m^-,\gamma_m^+\right]>$ as

$<\gamma_1,\gamma_2,\cdots,\gamma_m>$, and

$<\left[z_1^-,z_1^+\right],\left[z_2^-,z_2^+\right],\cdots,\left[z_m^-,z_m^+\right]>$ as

$<z_1,z_2,\cdots,z_m>$.

Definition 3.7. If every vertex in a CIVmPF graph $G=(V,W,F)$ has the same degree $<\delta_1,\delta_2,\cdots,\delta_m>$, then $G=(V,W,F)$ is called *regular*CIVmPFgraph or CIVmPFgraph of degree $<\delta_1,\delta_2,\cdots,\delta_m>$

Definition 3.8. If every edge in a CIVmPF graph $G=(V,W,F)$ has the same degree $<\delta_1,\delta_2,\cdots,\delta_m>$, then $G=(V,W,F)$ is called an *edge regular* CIVmPF graph.

Definition 3.9. If every edge in a CIVmPF $G=(V,W,F)$ has the same total degree $<\delta_1,\delta_2,\cdots,\delta_m>$, then $G=(V,W,F)$ is called *totally edge regular* CIVmPF graph.

Theorem 3.1.*Let* $G=(V,W,F)$ *be a CIVmPF graph on a cycle* $G^*=(V,E)$. *Then* $\sum_{v_j\in V}d_G(v_j)=\sum_{(v_j,v_k)\in E\ j\neq k}d_G(v_j,v_k)$.

Proof. Suppose that $G=(V,W,F)$ is aCIVmPF graph and G^* be a cycle $v_1v_2v_3\cdots v_nv_1$.

Now, we get

$$\sum_{j=1}^{n} d_G\left(v_j, v_{j+1}\right) = d_G\left(v_1, v_2\right) + d_G\left(v_2, v_3\right) + \cdots + d_G\left(v_n, v_1\right)$$

where $v_{n+1} = v_1$.

$$= d_G\left(v_1\right) + d_G\left(v_2\right) - 2\overline{NE}_{F(v_1, v_2)} +$$

$$d_G\left(v_2\right) + d_G\left(v_3\right) - 2\overline{NE}_{F(v_2, v_3)} + \cdots$$

$$+ d_G\left(v_n\right) + d_G\left(v_1\right) - 2\overline{NE}_{F(v_n, v_1)}$$

$$= 2\sum_{v_j \in V} d_G\left(v_j\right) - 2\sum_{j=1}^{n} \overline{NE}_{F(v_j, v_{j+1})}$$

$$= \sum_{v_j \in V} d_G\left(v_j\right) + \sum_{v_j \in V} d_G\left(v_j\right) - 2\sum_{j=1}^{n} \overline{NE}_{F(v_j, v_{j+1})}$$

$$= \sum_{v_j \in V} d_G\left(v_j\right) + 2\sum_{j=1}^{n} \overline{NE}_{F(v_j, v_{j+1})} \quad - 2\sum_{j=1}^{n} \overline{NE}_{F(v_j, v_{j+1})}$$

$$= \sum_{v_j \in V} d_G\left(v_j\right).$$

Hence, $\sum_{v_j \in V} d_G\left(v_j\right) = \sum_{\left(v_j, v_k\right) \in E \; j \neq k} d_G\left(v_j, v_k\right)$.

Remark 3.1. Let $G = (V, W, F)$ be a CIVmPFgraph on a crisp graph G^*. Then

$$\sum_{\left(v_j, v_k\right) \in E} d_G\left(v_j, v_k\right) = \sum_{\left(v_j, v_k\right) \in E} d_{G^*}\left(v_j, v_k\right) \overline{NE}_{F(v_j, v_k)},$$

where $d_{G^*}\left(v_j, v_k\right) = d_{G^*}\left(v_j\right) + d_{G^*}\left(v_k\right) - 2$, for all $\left(v_j, v_k\right) \in E$.

Theorem 3.2. *Let $G = (V, W, F)$ be a CIVmPFgraph on a c-regular crisp graph G^*. Then* $\sum_{\left(v_j, v_k\right) \in E} d_G\left(v_j, v_k\right) = (c-1) \sum_{v_j \in V} d_G\left(v_j\right)$.

Proof. From Remark 3.1., we have

$$\sum_{\left(v_j, v_k\right) \in E} d_G\left(v_j, v_k\right) = \sum_{\left(v_j, v_k\right) \in E} d_{G^*}\left(v_j, v_k\right) \overline{NE}_{F(v_j, v_k)}$$

$$= \sum_{v_j v_k \in E} \left(d_{G^*}\left(v_j\right) + d_{G^*}\left(v_k\right) - 2\right) \overline{NE}_{F(v_j, v_k)}$$. Since G^* is a regular crisp graph, we have the degree of every vertex in G^* is c. i.e. $d_{G^*}\left(v_j\right) = c$, So

$$\sum_{\left(v_j, v_k\right) \in E} d_G\left(v_j, v_k\right) = \left(c + c - 2\right) \sum_{\left(v_j, v_k\right) \in E} \overline{NE}_{F(v_j, v_k)},$$

$$\sum_{\left(v_j, v_k\right) \in E} d_G\left(v_j, v_k\right) = 2\left(c - 1\right) \sum_{\left(v_j, v_k\right) \in E} \overline{NE}_{F(v_j, v_k)},$$

$$\sum_{\left(v_j, v_k\right) \in E} d_G\left(v_j, v_k\right) = (c-1) \sum_{v_j \in V} \left(d_G\left(v_j\right)\right).$$

Theorem 3.3. *Let $G = (V, W, F)$ be a CIVmPFgraph on a crisp graph G^*. Then*

$$\sum_{\left(v_j, v_k\right) \in E} td_G\left(v_j, v_k\right) = \sum_{\left(v_j, v_k\right) \in E} d_{G^*}\left(v_j, v_k\right) \overline{NE}_{F(v_j, v_k)} +$$

$$+ \sum_{\left(v_j, v_k\right) \in E} \overline{NE}_{F(v_j, v_k)}.$$

Proof. From the definition of total edge degree of G, we get

$$\sum_{\left(v_j, v_k\right) \in E} td_G\left(v_j, v_k\right) = \sum_{\left(v_j, v_k\right) \in E} \left(d_G\left(v_j, v_k\right) + \overline{NE}_{F(v_j, v_k)}\right)$$

$$= \left(\sum_{\left(v_j, v_k\right) \in E} d_G\left(v_j, v_k\right) + \sum_{\left(v_j, v_k\right) \in E} \overline{NE}_{F(v_j, v_k)}\right).$$

From Remark 3.1., we have

$$\sum_{(v_j,v_k)\in E} td_G(v_j,v_k) = \sum_{(v_j,v_k)\in E} d_{G^*}(v_j,v_k)\overline{NE}_{F(v_j,v_k)} + + \sum_{(v_j,v_k)\in E} \overline{NE}_{F(v_j,v_k)}.$$

Theorem 3.4. *Let $G=(V,W,F)$ be a CIVmPFgraph. Then $\overline{NE}_F$ is a constant function if and only if the following conditions are equivalent*

i) G is an edge regular CIVmPFgraph.

ii) G is a totally edge regular CIVmPFgraph.

Proof. Suppose that $\overline{NE}_F$ is a constant function.

Then $\overline{NE}_{F(\alpha,\beta)} = <\gamma_1,\gamma_2,...,\gamma_m> \forall (\alpha,\beta)\in E$.

Let G be an edge regular CIVmPFgraph. Then for all $(v_j,v_\gamma)\in E$, $d_G(v_j,v_\gamma) = <\delta_1,\delta_2,\cdots,\delta_m>$.

Now we have to show that G is a totally edge regular CIVmPFgraph.

Now $td_G(v_j,v_\gamma) = d_G(v_j,v_\gamma) + \overline{NE}_{F(v_j,v_\gamma)}$

$= <\delta_1,\delta_2,\cdots,\delta_m> + <\gamma_1,\gamma_2,...,\gamma_m>$

$= <\delta_1+\gamma_1,\delta_2+\gamma_2,\cdots,\delta_m+\gamma_m>$ for all $(v_j,v_\gamma)\in E$. Then G is a totally edge regular CIVmPF graph. Now, let G be a $<h_1,h_2,\cdots,h_m>-$ totally edge regular CIVmPFgraph.

Then $td_G(v_j,v_\gamma) = <h_1,h_2,\cdots,h_m>$ for all $(v_j,v_\gamma)\in E$.

So, we have

$td_G(v_j,v_\gamma) = d_G(v_j,v_\gamma) + \overline{NE}_{F(v_j,v_\gamma)} = <h_1,h_2,\cdots,h_m>$.

Hence $d_G(v_j,v_\gamma) = <h_1,h_2,\cdots,h_m> - \overline{NE}_{F(v_j,v_\gamma)}$

$= <h_1-\gamma_1,h_2-\gamma_2,\cdots,h_m-\gamma_m>$.

Then G is an $<h_1-\gamma_1,h_2-\gamma_2,\cdots,h_m-\gamma_m>-$ edge regular CIVmPFgraph.

Conversely, suppose that G is an edge regular CIVmPFgraph and G is a totally edge regular CIVmPF graph which are equivalent. We have to prove that $\overline{NE}_F$ is a constant function. In a contrary way, we suppose that $\overline{NE}_F$ is not a constant function. Then $\overline{NE}_{F(v_j,v_\gamma)} \neq \overline{NE}_{F(v_\delta,v_s)}$ for at least one pair of edges $(v_j,v_\gamma),(v_\delta,v_s)\in E$. Let G be an $<\delta_1,\delta_2,\cdots,\delta_m>-$ edge regular CIVmPFgraph. Then $d_G(v_j,v_\gamma) = d_G(v_\delta,v_s) = <\delta_1,\delta_2,\cdots,\delta_m>$.

Hence for every $(v_j,v_\gamma)\in E$ and for every $(v_\delta,v_s)\in E$;

$td_G(v_j,v_\gamma) = d_G(v_j,v_\gamma) + \overline{NE}_{F(v_j,v_\gamma)}$

$= <\delta_1,\delta_2,\cdots,\delta_m> + \overline{NE}_{F(v_j,v_\gamma)}$

$td_G(v_\delta, v_s) = d_G(v_\delta, v_s) + \overrightarrow{NE}_{F(v_\delta, v_s)}$

$= <\delta_1, \delta_2, \cdots, \delta_m> + \overrightarrow{NE}_{F(v_\delta, v_s)}$. Since

$\overrightarrow{NE}_{F(v_j, v_\gamma)} \neq \overrightarrow{NE}_{F(v_\delta, v_s)}$,

we have $td_G(v_j, v_\gamma) \neq td_G(v_\delta, v_s)$. Hence G is not a totally edge regular CIVmPFgraph. This is a contradiction to our assumption. Hence $\overrightarrow{NE}_F$ is a constant function. In the same way, we can prove that $\overrightarrow{NE}_F$ is a constant function, when G is a totally edge regular CIVmPFgraph.

Theorem 3.5 *Let G^* be a h-regular crisp graph and $G = (V, W, F)$ be aCIVmPFgraph on G^*. Then, $\overrightarrow{NE}_F$ is a constant function if and only if G is both regular CIVmPFgraph and totally edge regular CIVmPF graph.*

Proof. Let $G = (V, W, F)$ be aCIVmPFgraph on G^* and let G^* be a $h-$regular crisp graph. Assume that $\overrightarrow{NE}_F$ is a constant function. Then $\overrightarrow{NE}_{F(\alpha,\beta)} = <\gamma_1, \gamma_2, \cdots, \gamma_m> \forall (\alpha, \beta) \in E$ where $\gamma_1, \gamma_2, \cdots, \gamma_m$ are constants. From the definition of degree of a vertex, we get

$d_G(v_j) = \sum_{(v_j, v_\gamma) \in E} \overrightarrow{NE}_{F(v_j, v_\gamma)}$

$= \sum_{(v_j, v_\gamma) \in E} <\gamma_1, \gamma_2, \cdots, \gamma_m>$

$= <h\gamma_1, h\gamma_2, \cdots, h\gamma_m>$

for every $v_j \in V$.

So $d_G(v_j) = <h\gamma_1, h\gamma_2, \cdots, h\gamma_m>$ for every $v_j \in V$. Therefore, G is a regular CIVmPFgraph.

Now $td_G(v_j, v_\gamma) = d_G(v_j) + d_G(v_\gamma) - \overrightarrow{NE}_{F(v_j, v_\gamma)}$

$= \sum \overrightarrow{NE}_{F(v_j, v_z)} + \sum \overrightarrow{NE}_{F(v_\gamma, v_z)} - \overrightarrow{NE}_{F(v_j, v_\gamma)}$

$= \sum_{\substack{(v_j, v_z) \in E \\ j \neq z}} <\gamma_1, \gamma_2, \cdots, \gamma_m>$

$+ \sum_{\substack{(v_\gamma, v_z) \in E \\ \gamma \neq z}} <\gamma_1, \gamma_2, \cdots, \gamma_m>$

$- <\gamma_1, \gamma_2, \cdots, \gamma_m>$

$= h <\gamma_1, \gamma_2, \cdots, \gamma_m>$

$+ h <\gamma_1, \gamma_2, \cdots, \gamma_m>$

$- <\gamma_1, \gamma_2, \cdots, \gamma_m>$

$= (2h-1) <\gamma_1, \gamma_2, \cdots, \gamma_m>$

For all $(v_j, v_\gamma) \in E$.

Hence G is also a totally edge regular CIVmPF graph.

Conversely, assume that G is both regular and totally edge regular CIVmPFgraph.

Now we have to prove that $\overrightarrow{NE}_F$ is a constant function. Since G is regular, $d_G(v_j) = <z_1, z_2, \cdots, z_m>$ for all $v_j \in V$. Also G is totally edge regular.

Hence $td_G(v_j, v_\gamma) = <h_1, h_2, \cdots, h_m>$ for all $(v_j, v_\gamma) \in E$.

From the definition of totally edge degree, we get

$td_G(v_j, v_\gamma) = d_G(v_j) + d_G(v_\gamma) - \overline{NE}_{F(v_j,v_\gamma)}$ for all $(v_j, v_\gamma) \in E$.

$< h_1, h_2, \cdots, h_m > = < z_1, z_2, \cdots, z_m > + < z_1, z_2, \cdots, z_m > - \overline{NE}_{F(v_j,v_\gamma)}$,

so $\overline{NE}_{F(v_j,v_\gamma)} = 2 < z_1, z_2, \cdots, z_m > - < h_1, h_2, \cdots, h_m >$

$=< 2z_1 - h_1, 2z_2 - h_2, \cdots, 2z_m - h_m >$ for all $(v_j, v_\gamma) \in E$. Hence $\overline{NE}_F$ is a constant function.

4. Conclusions

In this article, order and size of a CIVmPF graph are defined. Further, equivalence condition for edge regular CIVmPFgraph and totally edge regular CIVmPF graph is proved. We will extend our work on CIVmPF planar graphs and find some of its properties along with the applications in image segmentation and image processing.

References

[1] *Akram M, Dudek W A (2011) Interval-valued fuzzy graphs. Computers and Mathematics with Applications 61: 289-299.*

[2] Akram M, Yousaf M M and Dudek W A (2015) Self centered Interval-valued fuzzy graphs. Afr. Mat 26(5): 887-898.

[3] Akram M (2012) Interval-valued fuzzy line graphs. Neural Computing Applications 21: 145-150.

[4] Bhutani K R and BatouAbdella (2003) On M-strong fuzzy graphs Information Sciences. 155: 103-109.

[5] Bhutani K R and Rosenfeld a (2003) Strong arcs in fuzzy graph. Inf. Sci. 152: 319-322.

[6] Chen J, Li S, Ma S, and Wang X (2014) m-polar fuzzy sets: an extension of bipolar fuzzy sets. Hindwai Publishing Corporation, the Scientific World Journal 2014: 1-8.

[7] Ghorai G and Pal M (2016) On some operations and density of m-polar fuzzy graphs. Pac. Sci. Rev. A Nath. Sci. Eng 17(1): 14-22.

[8] Ghorai G and Pal M (2016) Some operations of m-polar fuzzy graphs. Pac. Sci. Rev. A Nath. Sci. Eng 18(1): 38-46.

[9] Ghorai G and Pal M (2016) A study on m-polar fuzzy planar graphs. Int. J. Comput. Sci. Math. 7(3): 283-292.

[10] Ghorai G and Pal M (2016) Faces and dual of m-polar fuzzy planar graphs. J. Intell. Fuzzy Syst. 31(3): 2043-2049.

[11] Ghorai G and Pal M (2016) Planarity in vague graphs with application. ActaMathematicaAcademiaePaedagogiaceNyregyhziensis 33(2): 147-164.

[12] Ghorai G and Pal M (2016) Some isomorphic properties of m-polar fuzzy graphs with applications. Springer International Publishing 5(1): 2104-2125.

[13] Kaufmann. A (1975) Introduction to the theory of fuzzy Subsets. Academic Press, New York 1: 402-403.

[14] Mordeson N J and Nair S P (2000) Fuzzy graph and fuzzy hypergraphs. Physica-Verlag Heidelberg.

[15] Mordeson J N and Nair P S (1994) Operation on fuzzy graphs. Inf. Sci. 79(3-4): 159-170.

[16] Mandal S, Sahoo S, Ghorai G and Pal M (2017) Genus value of m-polar fuzzy graphs. Journal of Intelligent and Fuzzy Systems 34(3): 1947-1957.

[17] RamprasadCh, Varma PLN, Satyanarayana S, Srinivasarao N (2017) Vertex degrees and isomorphic properties in

complement of an m- polar fuzzy graph. Advances in Fuzzy Systems 2017: 1-9.

[18] RamprasadCh, Varma PLN, Satyanarayana S, Srinivasarao N (2017) Morphism of m-polar fuzzy graphs. Advances in Fuzzy Systems 2017: 1-9.

[19] Rashmanlou H, Samanta S, Pal M and Borzooei A R (2015) a study on bipolar fuzzy graps, Journal of Intelligent and Fuzzy Sustems 28(2): 571-580.

[20] Rashmanlou H, Samanta S, Pal M and Borzooei A R (2015) Bipolar fuzzy graphs with categorical properties, International Journal of Computational Intelligence Systems 8(5): 808-818.

[21] Rosenfeld A (1975) Fuzzy Graphs, Fuzzy sets and their Application Academic Press, New York. pp.77-95.

[22] Samanta S and Pal M (2015) Fuzzy planar graph. IEEE Transaction on Fuzzy Systems 23: 1936 – 1942.

[23] Sahoo S and Pal M (2016) Intuitionistic fuzzy tolerance graph with application Journal of Applied Mathematics and Computing 55: 495-511.

[24] Sahoo S and Pal M (2016) Intuitionistic fuzzy graphs and degree. Journal of Intelligent and Fuzzy Systems 32(1): 1059-1067.

[25] Sunitha S M and Mathew S (2013) Fuzzy graph theory: A survey. Ann Pure Appl Math 4: 92-110.

[26] Talebi A. A. And Rashmanlou H (2014) Complement and isomorphism on bipolar fuzzy graphs. Fuzzy Information and Engineering 6: 505-522.

[27] Talebi A. A. And Rashmanlou H (2013) Isomorphism on interval-valued fuzzy graphs annals of fuzzy Mathematics and Informatics 6(1): 47-58.

[28] Zadeh L A (1965) Fuzzy sets, Inf. Control: 338-353.

[29] Zadeh L A (1965) The concept of a Linguistic and application to approximate reasoning. Inf. Sci. : 199-249.

[30] Zhang. W. R,"Bipolar fuzzy sets and relations: a computational framework for cognitive modeling and multi agent decision analysis," (1994), Proceedings of IEEE Conf., 305-309.

[31] Zhang. W. R, "Bipolar fuzzy sets," (1998), Proceedings of FUZZY-IEEE Conf., 835-840.

REAL-LIFE APPLICATIONS OF ORDINARY DIFFERENTIAL EQUATIONS

V.Sankar Rao[1], Y.Rani[2]

1.V.Sankar Rao[1],Department of Mathematics, Malla Reddy College of Engineering,Hyderabad-500100
Email: sankar.achiever@gmail.com
2. Y. Rani[2],Department of Mathematics, Malla Reddy College of Engineering, Hyderabad-500100
Email: raniyrani88@gmail.com

ABSTRACT:

This study introduces real-life mathematical models of international relationships suitable for ordinary differential equations, by investigating conflicts between different nations or alliances. The system of differential equations are constructedbased on the work of Richardson. The solutions and the stability of systems of Oridinary Differential Equations are observed.

KEY WORDS: ODE, Arms race, Eigen values, Critical point, Armaments.

INTRODUCTION:

Ordinary differential equations (ODEs), especially systems of ODEs, have been applied in many fields such as physics, electronic engineering and population dynamics. This is a powerful tool for analysing the relationship between various dynamic quantities.In this paper we focus on the study of models describing international conflicts, which were originally derived by Lewis Fry Richardson, describing the relationship between two nations or two alliances that deem war to be imminent. In particular, he devised mathematical models of arms races using differential equations. One assumes that if one country increases its weapons, another country will do the same. Sequentially, the first country responds by storing more weapons. Richardson proposed that this kind of arms race can be represented by a pair of differential equations. Richardson's model of international relations, which includes an arm race, used for discussing stability, is analogous to the differential equations in the predator-prey model. In the present paper, firstly, a simpler model of an arms race is depicted. More realistic models are constructed, and additional factors that influence the relationship between two nations or alliances, such as the cost of armaments, the grievances between nations and their ambitions, are considered.

1. SIMPLE ARMS RACE FOR DIFFERENTIAL EQUATIONS:

It is a well-known fact that an increase in armaments is one of the primary reasons for war. Another reason is the unsolvable conflict of ambitions, such as occupying more territory or recovering tracts of land.

We assume that if one nation increases its armaments, then the opposing nation will do likewise because it assumes that the balance of power will be negatively affected.

Let x(t) be the armaments of nation X, and y(t) be the armaments of nation Y at time t. The rate of change of the armaments on one side depends on the number of armaments on the opposing side, because if one nation increases its armaments, the other will follow suit. That is, $dx/dt\ (or\ dy/dt)\ is\ proportional\ to\ y\ (or\ x)$.We assign constants of proportionally h and 1 to x and y, respectively, which represent the efficiency of increasing armaments.
Hence, we can establish a system of differential equations in the following form:

$$\frac{dx}{dt} = kx \qquad (1.1)$$

$$\frac{dy}{dt} = ky$$

Thissystemcanbeusedtodescribetherelationshipbetweentwonationsoralliances,eachofwhichdecidestodefenditselfagainstpossibleattackbytheother.

It is easy to obtain the solutions for the system (1.1) which we give as follows:

$$x(t) = \sqrt{\frac{k}{l}}\ (Ae^{t\sqrt{kl}} - Be^{-t\sqrt{kl}}),$$

$$y(t) = (Ae^{t\sqrt{kl}} - Be^{-t\sqrt{kl}}) \qquad (1.2)$$

Given initial conditions,
x (0)=x_O,y (0)=y_O
we can obtain

$$A = \frac{1}{2}\left(y_0 + \sqrt{\frac{l}{k}}x_0\right) \quad (1.3)$$

$$B = \frac{1}{2}\left(y_0 - \sqrt{\frac{l}{k}}x_0\right)$$

It is possible to estimate the values of $k\ and\ l$. For example, when y remains a constant C, it follows from (1.1)that

$$(1.4)\ \frac{1}{k} = \frac{C}{\frac{dx}{dt}} = C\ \frac{dt}{dx}$$

Solving (1.4) we obtain

$$(1.5) \qquad \frac{1}{k}x = Ct + b$$

Assuming x(0) = 0, it follows from (2.5)

that b = 0and

$$\frac{1}{k} = \frac{C}{x}t\ \ for\ x > 0$$

Hence when Xhas caught up to Y, which means X= C, we have$\frac{1}{k} = t$ Thus $\frac{1}{k}$ is the time required for nation Xto catch up with the armaments of Y provided that y remains constant. Richardson also observed that k is proportional to the amount of industry in a country.

2. CONSTRUCTION OF A REALISTIC MODEL

The relationship between nations or alliances in the real world is more complicated. We therefore need to modify the system (2.1) by considering more factors thataffectthechangeratesdx/dtanddy/dtinanefforttoadaptittotherealworld. Richardsonconstructeddifferentialequationsofconflict,takingintoaccountfactors suchasthecostofarmaments,grievancesorambitionsbetweennations,etc([1], 1993;[11],1957).Thesystemconstructedfordescribingtherelationshipbetween the nations or alliances, X and Y, is asfollows:

$$\frac{dx}{dt} = ky - \alpha x + g \qquad (2.1)$$

$$\frac{dy}{dt} = lx - \beta y + h, \qquad k, l, \alpha, \beta, g, h > 0$$

where $x(t)$ $(respectively\ y(t))$ denotes the armaments of nation X(Y); $k(l)$ is the efficiency of increasing the armaments of X(Y); $g(h)$ is the ambitions of or grievances that X(Y) has towards Y(X) , affecting dx/dt (dy/dt) positively the influence of the cost of armaments is a restraining factor, represented by $-\alpha x(\beta y)$.

We encourage students to observe the stability of system (1.1) by finding Σ and analysing the critical point. This is an excellent exercise for students to familiarise themselves with the classification of a critical point by stability and type, utilising the eigenvalues. The stability of the critical point of the system (2.1) depends on the logical relationship between $kl\ and\ \alpha\beta$.

It would entail a great deal of work to simplify the solutions to system (2.1).

Given the initial conditions $x(0) = x_0, y(0) = y_0$, the unique solution to (2.1) can be written as follows:

$$x(t) = x_1 + \frac{kl}{2\mu(\mu+\omega)}\ Ae^{\lambda_1 t} + \frac{1}{2}\frac{\mu+\omega}{\mu}\ Be^{\lambda_2 t} \quad (2.2)$$

$$y(t) = y_1 + \frac{l}{2\mu}\ Ae^{\lambda_1 t} - \frac{l}{2\mu}\ Be^{\lambda_2 t} \quad (2.3)$$

Where $(x_1, y_1) = \left(\frac{\beta g + hk}{\alpha\beta - kl}, \frac{\alpha h + gl}{\alpha\beta - kl}\right)$
Is the critical point

$$\lambda_1 = -\frac{1}{2}(\alpha+\beta) + \frac{1}{2}\sqrt{(\alpha-\beta)^2 + 4kl}$$

$$\lambda_2 = -\frac{1}{2}(\alpha+\beta) - \frac{1}{2}\sqrt{(\alpha-\beta)^2 + 4kl}$$

Are Eigen values and

$$\omega = \tfrac{1}{2}(\alpha-\beta), \mu = \sqrt{\omega^2 + kl} \quad (2.4)$$

$$A = \left(x_0 + \frac{g}{\lambda_1}\right) + \frac{\mu+\omega}{l}\ (y_0 + \frac{h}{\lambda_1}) \quad (2.5)$$

$$B = \left(x_0 + \frac{g}{\lambda_2}\right) - \frac{k}{\mu+\omega}\ (y_0 + \frac{h}{\lambda_2}) \quad (2.6)$$

Finally, we estimate the coefficients in our model. It is interesting to note that Richardson estimates $\alpha^{-1}(\beta^{-1})$ to be thelifetime of X's (Y's) parliament ([2]). For example, since the lifetime of Britain's parliament is five years, we obtain α= 0.2 for that country.

To estimate $k\ and\ l$, for example, we consider $g = 0\ and\ y = C$. Hence

$$\frac{dx}{dt} = kc - \alpha x$$

Assuming $x(0) = 0$ and solving the above equation, we obtain:

$$x = \frac{kC}{\alpha}\ (1 - e^{-\alpha t}) \quad (2.7)$$

Substituting the power series expansion

$$e^{-\alpha t} = \sum_{n=0}^{\infty} \frac{(-\alpha t)^n}{n!} = 1 - \alpha t + \cdots \ldots \ldots \ldots$$

In (2.7), since $\alpha t < 1$, $we\ have$

$$x \approx \frac{kC}{\alpha}\ \alpha t \quad for \quad t > 0$$

That is $\frac{1}{k} = \frac{C}{x}\ t$

Thus $\frac{1}{k}$ isthetimerequiredforX tocatchuptoY,ifY'sarmamentsonlyremain a constant C. We recall that k represents the product efficiency of armaments of nationX.Itisobviousthatausefulexerciseistoapp lytheknowledgetoseries.

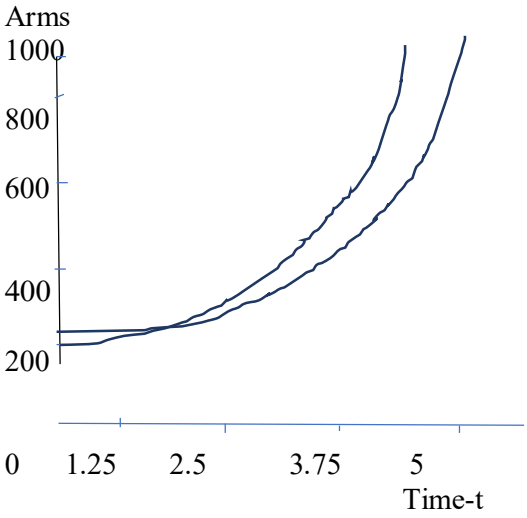


FIGURE1:Solutionstothemodel(3.1)with k=0.6,l=0.8,$\alpha = \beta$=0.2, $x_0 = 100$, $y_0 = 80$.

expansion of functions. We note that our method of estimating k is different from the method given in [2], in which the constant k is estimated in the following way, assuming $g = 0$ *and* $y = y_1$:
when $x = 0$,

$$\frac{1}{k} = \frac{y_1}{\frac{dx}{dt}}.$$

A future work could include the collection of data and information on historical wars to construct mathematical models. For example, during the Cold War, both the USA and the Soviet Union were involved in an arms race for conventional and nuclear weapons.

Here we only give some examples and graphs, choosing different values for the constants k, l, α, β. We assume $g = h = 0$, so $x_1 = y_1 = 0$. We use a solid line to represent the solution $x(t)$, and a grey line for the solution $y(t)$.

If k = 0.6, l = 0.8, $\alpha = \beta = 0.2$, $x_0 = 100$, y_0=80, the solutions and graphs are as follows (unstable case, $\alpha\beta < kl, \lambda_1 > 0 > \lambda_2$) (see Figure 1)

$$x(t) = 84.641\ exp\ (0.49282t) + 15.359\ exp\ (-0.89282t)$$

$$y(t) = 97.735\ exp\ (049288t) - 17.735\ exp\ (-0.89282t)$$

If k = 0.35, l = 0.4, $\alpha = 0.3, \beta = 0.5$, $x_0 = 100$, y_0=95,

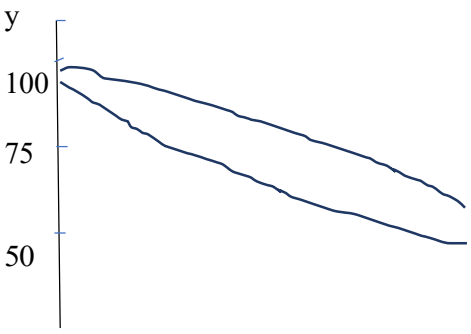
the solutions and graphs are as follows (unstable case, $\alpha\beta < kl, \lambda_2 < \lambda_1 < 0$) (see Figure 1)

$$x(t) = 105.84\ exp\ (-1.2702 * 10^{-2}t) - 5.8355\ exp\ (-0.7873t)$$

$$y(t) = 8.1247\ exp\ (-0.7873t) + 86.875\ exp\ (-1.2702 * 10^{-2}t)$$

Conclusion:

We believe that more research could be conducted into the mathematical modelling of international relationships at undergraduate level. Such modelling also has realistic applications in military, business and other fields. From an educational perspective, these mathematical models are also realistic applications of ordinary differential equations (ODEs) — hence the proposal that these models should be added to ODE textbooks as flexible and vivid examples to illustrate and study differential equations.



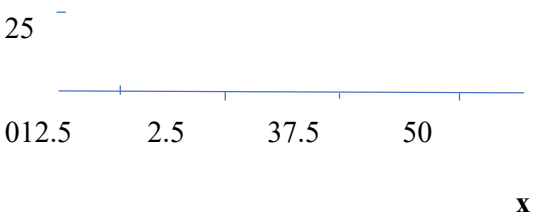


Figure2: Solutions to the model (3.1) $\boldsymbol{with\ k = 0.6,\ l = 0.8\ \alpha = \beta = 0.2, x_0 = 100, y_0 = 80}$.

REFERENCES:

[1] *ASHFORD, O.M. 1993. Collected papers of Lewis Fry Richardson. Cambridge.*

[2]DENNIS,G.Z.&MICHAEL,R.C.1997.Differential equationswithboundaryvaluepvoblems. Brooks &Cole.

[3]JOUBERT,S.V.2002.Technicon MathematicsIIIwithDERIVE5.GreatWhitePublishers,Faerie Glen,Ganteng,SouthAfrica.

[4]KREYSZIG,E.1999.AdvancedEngineering mathematics.8thed.NewYork:Wiley&Sons.

[5]LANCHESTER, F.W. 1996. Aircraft in warfare: the down of the fourth arm. London: Constable.

[6]RICHARDSON,L.F. 1960a.Armsandinsecurity:amathematicalstudyoftheCAUSESandoriginsofwar. EditedbyN.Rashesky&E.Trucco.Pittsburgh:BoxwoodPress..

GET OCEAN OF VOCABULARY, GET COMMAND OVER THE LANGUAGE

Name of Author 1: Humera Nafees
Assistant Professor
Malla Reddy College of Engineering
hum.nafees@gmail.com
+91808374171

Name of the Author 2: K.Swathi
Assistant Professor
Malla Reddy College of Engineering
Swathi4645@gmail.com
+918374461994

KEY WORDS:

Abstract:
This paper is to enlighten the importance of Vocabulary in language learning. Learning is a continuous process; it needs a full time effort to learn anything. When it comes to vocabulary learning, it needs curiosity, interest, hard work and consistency. Learning 1st language or native language learning is easy and fast compare to new language learning depend on personal interest or need of the particular language learning.
To make oneself good and fluency speaker; vocabulary plays a vital role. This vocabulary building can do via reading listening and speaking. The more she/he practice and research the more she/he will get command over the language. Various techniques are useful for vocabulary building it plays the 'key role' for getting command on new language or foreign language.

I. INTRODUCTION:

Human being's learning starts soon after their birth, 1st they will learn gesture, speaking tone, love with affection and facial expression. Gradually they learn to respond and show their feel or needs, even though they can't talk but by their cry smile and laugh they will make others understand what and how they feel and about their needs.
This learning leads to learn vocabulary building, as they have the ability of hearing and talking but don't know how to talk or how to response for what they heard—they slowly start making noise slimier to their speaker, and imitate the words, and the learning process starts. At beginning they can learn 50 to 60 words. It increases by time passes. Vocabulary learning is a continuous process, and follows the same rule for all languages, and needs extra efforts to learn other language than 1st language. Importance of vocabulary in all language, learning& utilizing the words correctly, suggestions to learn

What is Vocabulary?

- Vocabulary is the bank of words in different languages.
- Vocabulary is the necessary word bank; which need to memorise with meaning and usage in new language learning.
- Vocabulary is the bridge between the learning and language.

Types of Vocabulary:

There are four types of vocabularies, they are Listening, Writing, Reading and Speaking.

- **Listening Vocabulary:**
  This skill getting developed before the birth of human, babies inside the womb can listen the words. After the birth babies listen every words, gradually they start responding for 20% to 40% words they listen, this process is a

continuous process an adult can comprehend 80% to 90% words what she/he listen.

When it comes to dumb kid; sign language is the only option to learn new language here ratio is very low.

- **Writing Vocabulary:**
  Writing vocabulary is the words which human and writes by knowing its meaning and context. Here we use punctuation marks to express emotion or to give signal to the reader, as in writing we don't have an opportunity to express ourselves through facial expression or modulation of voice/tone. It has 3 levels.
  Level-1: here we start learning letter and joining them together. In this level easy spelling given to the learner which he/she can remember easily. Eh. Can, will, cat, bat, girl, brother etc also learn basic expressive words such as happy, sad, fast, help, slow etc. Class one to 7th students.
  Level-2: here learners will learn words with high frequency such as, empathy, condolence, etc. Class 8th to 12th students.
  Level-3: here learners will learn words from specific category such as Science, Maths, Geography and Medicine etc, class 12th to UG and PG students.
- **Reading Vocabulary:**
  It is second largest vocabulary bank, it is related to the text we read, we can read enumerable words but it is not necessary that we understand all, this skill is related to our comprehension that what we read we understand it. Very often we don't use the words, we read in our daily life limitation of this vocabulary is varying person to person and their need.
  In the beginning human can read and understand basic language, gradually they get the command over the language in result can read and comprehend complex text.
- **Speaking Vocabulary:**
  Speaking Vocabulary refers to the words we speak. Our speaking vocabulary is restricted. Most of the adults use almost 5000 to 10000 words for communication (for conversation and giving instructions). The number of words used in this case is far less than listening vocabulary, the reason being the level of comfort in usage.
  At early age we can speak few words slowly it increases and development t takes place at the age of adult and it get finer with experience and usage.

Category of Vocabulary:

a. **Receptive Vocabulary**
   Receptive vocabulary is words that learners recognize and understand when they are used in context, but which they cannot produce. It is vocabulary that learners recognize when they see or meet in reading text but do not use it in speaking and writing

b. **Productive Vocabulary**
   Productive vocabulary is the words that the learners understand and can pronounce correctly and uses constructively in speaking and writing. It involves what is needed for receptive vocabulary plus the ability to speak or write at the appropriate time. Therefore, productive vocabulary can be addressed as an active process, because the learners can produce the words to express their thoughts to others (ALQAHTANI, 2015)

II. TECHNIQUES TO DEVELOP VOCABULARY:

Learn and use New Words Every Day :-

To improve your vocabulary quickly, make an effort to learn at least one new word every

singleday. There are plenty of ways to do this, such as a Word of the Day calendar or email list, or simply picking a word from a thesaurus or dictionary. Once you learn a new word try to use those words in our daily life. Use the new words in your daily conversations, journal entry, an assignment or an email to a friend in this way you can have a common over the language.

Look up Words You Don't Know

we often come across new words that are unfamiliar to us. Don't just gloss over them but taketime to look them up, and if you don't have the time right then, write them down and look them up later in this way we learn new words.

III. AVOID THE TEMPTATION TO TRANSLATE WORDS INTO YOUR NATIVE LANGUAGE.

Instead, write a few synonyms or a sentence that makes the meaning of the word obvious. This will force you to think about the concepts and the meanings of the words instead of their translations in your native language.

Learn in a peaceful setting.

Block out any distractions: close your door, wear earplugs, or switch your phone to airplane mode. Set a timer for your session. It will help you to learn the new vocabulary very easily.

Make Your Own Vocabulary Tests

Keep a list of the new words you learn each week and incorporate into writing and conversation. At the end of each week, make yourself a quiz using the words to cement them in your memory.

Practice with Word Games

Playing word games is another simple tactic for improving the vocabulary. Start with a quick game that requires no tools other than your friends and your brain. For variety, players can take a slightly different route and suggest a synonym to the word. Players can also challenge the answer if they think someone is incorrect or bluffing. Referring to the Dictionary is great way to check answers and learn new words along the way. The point of the game is to bring an element of novelty and fun to learning and recalling a variety of words (how to increase your vocabulary, 2018)

Conclusion: - English vocabulary is complex , when learning vocabulary it needs lot of curiosity and personal interest . vocabulary plays a vital role in every person's life. Vocabulary learning is a continuous process and follows the same rules for all the languages. There are different types of vocabularies they are listening, writing , speaking and reading. vocabulary is words that learners recognize and understand when they are used in contextand uses constructively in speaking and writing. Vocabulary can be built by learning new words every day and by using them in daily conversations and get command over the language by using the proper vocabulary.

References:

ALQAHTANI, M. (2015). THE IMPORTANCE OF VOCABULARY IN LANGUAGE. International Journal of Teaching and Education Vol. III, No. 3 / 2015 .

how to increase your vocabulary. (2018). Retrieved from grammer.yourdictionary.com: https://grammar.yourdictionary.com/for-students-and-parents/how-to-increase-your-vocabulary.html

MOMENT OF INERTIA BASED RADIAL CODING FEATURES OF INVARIANT CHARACTER RECOGNITION USING FUZZY MIN-MAX NEURAL NETWORKS

1)R. Vijayasree, Asst.Prof
Mallareddy College of Engineering
Hyderabad.
rv.jayasree@gmail.com

2)P.Vijai Bhaskar, Professor
Geethanjali College of Engineering and Technology
Keesara (m), Cheeryal (v),
pvijaibhaskar@gmail.com

ABSTRACT

This paper proposes a character recognition system that is invariant to translation, rotation and scale. The system has two main components feature extraction and recognition. The feature extraction is carried out using moment of inertia based radial coding features. The main advantage of this feature vector is that it doesn't require the normalization of character. These features also easy to understand and implement compared to other methods computer requirements are also negligible. The Fuzzy Min-Max neural network (FMNN) is used in the recognition phase. The nine dimensional feature vector consists of Normalized moment of inertia and eight radial coding features. The character recognition systems is tested on 26 uppercase typed and hand written English letters. This character recognition system is also tested on different fonts (Ariel Unicode, Ariel Narrow, Microsoft scan serif) and hand written characters of five different writers.

Key words: Character Recognition System, Fuzzy Min-Max Neural network, Invariant Character Recognition, Moment of Inertia, and Radial Coding Features.

I. Introduction

Character recognition is one of the most successful applications of pattern recognition. Automated character recognition is of vital importance in many industries such as banking and shipping. A variety of character recognition methods are available such as boundary-based analysis via Fourier descriptors [3], neural-networks models [4] and invariant moment's [5]. Boundary-based analysis using discrete Fourier transforms has been proposed for character recognition. Algorithms based on this kind of analysis are called Fourier descriptors and basically, invariance is obtained by normalizing the frequency representation of the character shape. Its major drawback is that it is unable to cope with large translations and rotations in the character. High-order networks have been utilized recently for invariant recognition.

In this type of model, one has to encode the properties of invariance in the values of the synaptic weights. The relations between pixels of the characters are used, and the invariance is directly constructed in the network. The number of combinations of possible relations between pixels increases in a nonlinear proportion to the number of input data. This is the main disadvantage of this approach. However, most of these methods are too computationally expensive or are not invariant under the three types of transformations: scaling, translation and rotation.

Invariant Character Recognition (ICR) that achieves excellent invariance under translation, rotation and scaling is proposed. The main contribution in this paper is the development and implementation of one new feature and use different character recognition algorithm for the improving character recognition performance.

II. Basic Block Diagram of Recognition System

Recognition can be defined as class assignment for input pattern that are not identical to the patterns used for training of

the classifier. Figure 1 shows the basic block diagram of recognition or classification system [2]. The recognition system consists of input transducer providing input pattern data to feature extractor. Inputs to the feature extractor are sets of data vectors; each such set of data vector consists of real numbers for given application. The converted data at the output of the transducer can be compressed while still maintaining same level of machine performance called as features. The feature space dimensionality is to be much smaller than dimensionality of pattern space. Classifier assigns a class to input pattern by using extracted features of feature extractor.

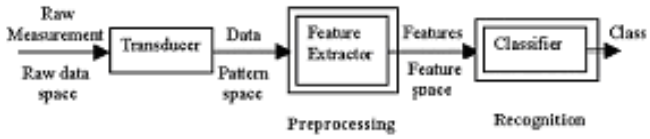


Figure 1 Basic block diagram of recognition system

III. PREPROCESSING

In invariant character recognition method, preprocessing is defined as the extraction of appropriate invariant features that are then used for recognition by a classification system. Feature can be defined as quantitative description of input character with in less dimension space and which is invariant under translation rotation and scale. The features play an important role in the recognition and classification system. Because the total information related to input character consists with in the extracted feature values. According to these feature values the high discriminate power of the classifier classifies the input character. The invariant features in this method are real numbers that are fed as vectors to the classification system.

The proposed invariant feature extraction takes reference from the centroid of the binary character. To find the centroid of two-dimensional object treat character as a very thin plate divide the object into small areas dA. For finding the centroid of object take the sum of the product of each area and the distance to an axis then divide by the total area of the object.

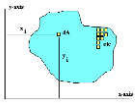


Figure.2 Centroid of the 2-D object

Centroid of character is calculated as follows:

For a given character the summation of product of each pixel x-coordinate and its gray value and divide the total number of character pixels in the character gives the x-coordinate of centroid.

$$C_x = \frac{\sum_{i=1}^{N} x_i * f(x,y)}{N} \qquad (1)$$

Similarly y-coordinate of centroid can be calculated.

$$C_y = \frac{\sum_{i=1}^{N} y_i * f(x,y)}{N} \qquad (2)$$

The centroid position is constant, even if the character is translated from its original position or rotates or scaled it. Taking reference as the centroid the Normalized moment of inertia radial coding features are found.

3.1 Normalized Moment of Inertia (NMI)

In general the moment of inertia quantifies the inertia of rotating object by considering its mass distribution. The moment of inertia is normally calculated by dividing the object into N-small pieces of mass m_1, m_2…..m_N. Each pieces at a distance r_1, r_2…r_N from the axis. The moment inertia of the object is

$$I = \sum_{i=1}^{N} d_i^{\,2} = \sum_{i=1}^{N} ((x_i - C_x)^2 + (y_i - C_y)^2) \qquad (3)$$

where C_x, C_y are the centroid co-ordinates. x_i, y_i the image pixel co-ordinates of the character. 'N' is the total number of pixels in the character. By dividing moment of inertia by N^2 (we will name it I_N) we get the Normalized moment of inertia. Due to the finite resolution of any digitized image, a rotated character may not conserve the number of pixels intact. So moment of inertia may vary but using normalized moment of inertia reduces this problem. The value of normalized moment of inertia invariant under translation rotation and scale invariant. The character 'A' and 'B' with different

orientations showed in Figure 3, and results shown in Table 1

(a) 0 (b) 25 (c) 55 (d) 75 (e) 90

(a) 0 (b) 25 (c) 55 (d) 75 (e) 90

Figure.3 Testing images of letter 'A' and 'B' for different orientations

Table 1: NMI Features for Characters 'A' and 'B' for different orientations.

Cases		(a)	(b)	(c)	(d)	(e)
NMI	A	0.2902	0.2933	0.2985	0.2940	0.2957
	B	0.3061	02768	0.2819	0.2760	0.2838

3.2 Radial Coding Features

The radial coding features are based on the fact that circle is the only geometrical shape that is naturally and perfectly invariant to rotation. In this we consider the number of intensity changes on two circular boundaries of some radius inside the character as it crosses it. This simple coding scheme extracts the topological characteristics of the character regardless of its position orientation and size. The methodology to obtain the radial coding features of a character can be summarized as follows:
1) Obtain the centroid of the character.

2) Generate *K* equidistant constant circles C_i around the centroid. The spacing is equal to the distance between the centroid and furthest pixel of the object divided by *K*.

3) For each circular boundary, count the number of intensity changes (zero to one or one to zero) that occur in the image.

In the above parameters the first value is the centroid pixel value. The remaining values are number of changes in pixel values on the circle. Combining these normalized moment of inertia and radial coding features we develop a feature vector call as moment of inertia based radial coding features is input to the classification system.

IV. CHARACTER RECOGNITION SYSTEM

Proposed block diagram of character recognition system consists of two main blocks Preprocessing and Recognition shown in Figure 4. In preprocessing block we extract the Normalized moment of inertia and radial coding features of character. The radial coding method gives the number of intensity changes on the circular boundaries of some radius inside the character around the centroid develops eight features. Combining these features to develop a ten dimensional training feature vector in which nine attributes are features tenth one assigned as pattern class. For testing the characters the feature vector is of nine dimensions only except the class of character.

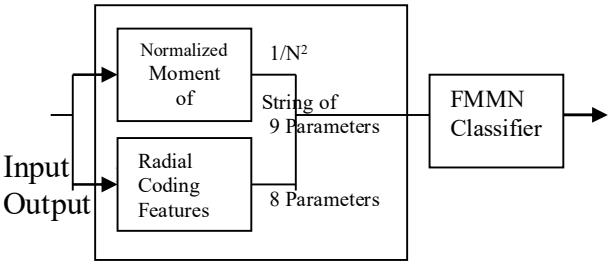


Figure 4 Block diagram of character recognition system.

The fuzzy min-max neural network (FMMN) proposed by Patrick Simpson [8] is a supervised

learning neural network (NN) classifier that utilizes fuzzy sets as pattern classes. A fuzzy set hyperbox is an n-dimensional box defined by a min point and a max point with corresponding membership function. The min-max points are determined using the fuzzy min-max learning algorithm. An expansion-contraction process learns nonlinear class boundaries in a single pass through data and provides the ability to incorporate new and refine existing classes without retraining.

It is a three layer feed forward neural network shown in Figure 5. F_A, F_B, and F_C represent these three layers respectively. F_A layer consists of *n* processing nodes equal to the dimension of the input pattern. The number of nodes in F_B layer is created during training, each F_B node in this layer represents a hyperbox fuzzy set where F_A to F_B connections is the min-max points of hyper box and the F_B transfer function is the hyper box membership function. The min points are stored in matrix *V* and max points are stored in matrix *W*. The connections are adjusted using learning algorithm. The F_C layer consists of *m* nodes, each F_C node represent a class.

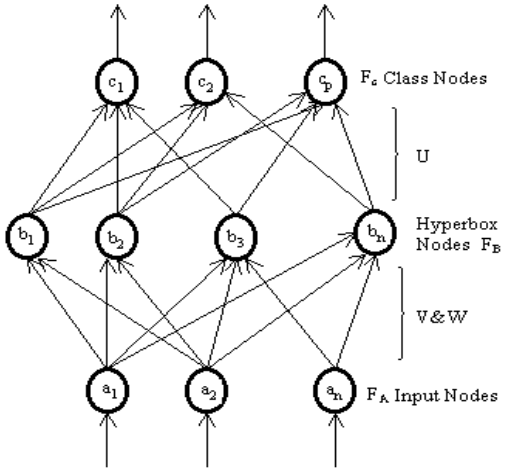


Figure.5 The Architecture of FMMN

The FMMN learning algorithm has four steps. They are initialization, hyperbox expansion, hyperbox overlap test and contraction. Initialization process is to create the first hyperbox with the first input pattern equal to the min-max points of the hyperbox. Hyperbox expansion processes test the expansion criteria of all ready created same class of max membership value hyperbox with presented input pattern. Overlap test allows the overlap between the hyperboxes from the same class and eliminates the overlap between the hyperboxes of different classes. Checks the overlap with the all dimensions of hyperbox and stores the minimum overlap dimension. In contraction process contract the overlapped hyperboxes along the minimum overlap dimension. The last three steps are repeated for all input patterns.

V. RESULTS

The character recognition system is tested using the 26 uppercase letters of the alphabet. *Twelve* different sizes and *eight* different orientations of each of 26 alphabets ninety-six experimental characters for each of the alphabet and in total 2496 experimental characters for all alphabets were generated. The largest character consists of 65x65 white pixels and smallest character consists of 15x15 white pixels. The Figure 6 shows few of the generated characters of letter E for different orientations and different sizes.

Learning Phase:

In order to obtain increased noise tolerance, during learning stage fifty percent of the randomly selected patterns from the total database used for learning. These patterns are called as *training set* including the class of character. Ten-dimension feature vector forms each training vector, in which first nine attributes are features and last one is its identifier. The first attribute is the normalized moment of inertia. The remaining eight features are number of changes of intensity pixel values, when the eight circles are intersected around the centroid of the character.

Recognition Phase:

Each one of the 96 characters for all the 26 letters were used as the *testing data set*. During this recognition phase all the extracted features of testing data set except class were presented to the character recognition system and their classes were found out. The results were then compared with original class of the character presented.

Case	NMI	CHARACTER 'X'
(a)	0.3381	1 0 3 4 4 4 4
(b)	0.3036	1 0 2 3 4 4 4
(c)	0.3146	1 0 2 4 4 4 4
(d)	0.3133	1 0 3 4 4 4 4
(e)	0.3356	1 0 3 4 4 4 4
(f)	0.3218	1 0 3 4 4 4 4
(g)	0.3384	1 0 3 4 4 4 4
(h)	0.3426	1 0 3 4 4 4 4

It is important to mention that most of the characters are used to test the method present a certain degree of noise or deformation. The noise is intrinsically produced during the transformation of the letters to other sizes and orientations.

Table 2 demonstrates moment of inertia based radial coding features of character 'X' with different sizes and different orientations (0, 10, 25, 30.55, 75, 90, and 110).

One can observe that the moment of inertia based radial coding features for character 'X' almost constant (with marginal variations) even though size, position and rotation of the character are changed.

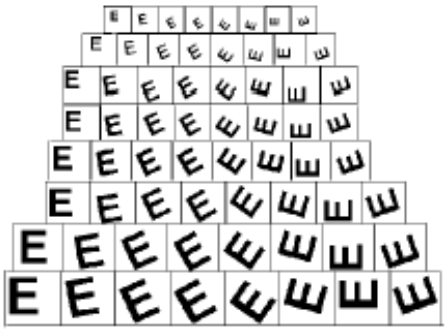
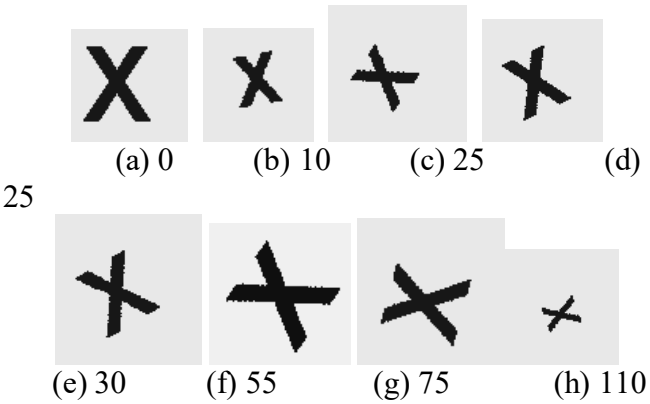


Figure 6: Testing images of letter 'E' for different orientations

Table 2: Moment of inertia based radial coding features for letter 'X' for different orientations and different sizes.



Figue 7: Testing character of letter 'X' for different orientations

These feature vectors can be used as input to the FMMN classification algorithm. The character recognition system is implemented on Pentium IV 1.4 GHz PC. The total database is used for the testing set of recognition purpose. The algorithm tested for this database moment of inertia based radial coding features. The percentage of recognition rate can be defined as

$$RECOGNITIONRATE = \frac{Number\ of\ tested\ samples\ being\ correctly\ recognized}{Total\ number\ of\ tested\ samples}$$

Initially, the FMMN algorithm is trained with different size of trained parameter θ *and* γ (sensitivity parameter) equal to one.

5.1 Results of single type of fonts

The results obtained with combination of moment of inertia and radial coding features shown in Table 3 and Table 4. The Table 3 shows the FMMN algorithm creates 404 hyperboxes and yields 100% recognition for training database. The table 4 shows the performance of FMMN algorithm with testing database and gives 98% recognition at which θ=0.03.

Table 3: Hyperboxes created and recognition rates obtained with the Training data set of moment inertia based radial coding Features

Font type	%Recognition rate
Microsoft Scan serif	78.0
Ariel Unicode	82.0
Ariel Narrow	78.0

Hyperbox size	No.of created	%Recognition
0.05	280	97.5
0.04	331	97.9
0.03	404	98.0

Table 4: Hyperboxes created and recognition rates obtained with the Testing data set of moment inertia based radial coding Features

Hyperbox size	No.of created	%Recognition
0.02	404	100

The trained classifier with moment of inertia based radial coding features used to find out the average percent recognition for each of the 26 letters. The ninety-six patterns of each character samples are given as input to the character recognition system

Letters	A	B	C	D	E	F	G	H	I	J	K	L	M
Percent Recognition	100	94	100	98	97	98	99	99	100	98	100	99	96

finds out the average percent recognition. The Table 5 shows the average percent recognition for each of the 26 letters and results shown in graph of Figure 8.

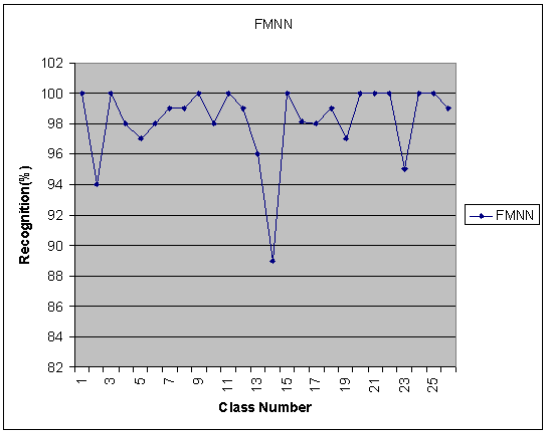


Figure 8 Average percent recognition of each of the 26 letters with graph.

Table 5: Average percent recognition for each of the 26 letters

Letters	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Percent Recognition	89	100	98	98	99	97	100	100	100	95	100	100	99

5.2 Results of different type of fonts

The character recognition system was also used test the different fonts (Arial Unicode, Arial Narrow, Microsoft scan Serif) of characters. The algorithm is trained with single font of database test with the different fonts. The Table 6 shows the percent recognition of different fonts. Mixed font of database prepared by taken a fifty percent of the data from each font of character. The algorithm is trained with mixed fonts of database test with the different fonts.

The Table 7 shows the percent recognition of different fonts.

Table 6: Percentage Recognition Rates of Three Different types of Font with Single font in Training at θ=0.03

Table 7: Percentage Recognition Rates of Three Different types of Font with Mixed fonts in Training at θ=0.03

Font type	%Recognition rate
Microsoft Scan	93.0
Ariel Unicode	97.0
Ariel Narrow	86.0

5.3 Results of handwritten characters

Character recognition system was also used to test for hand written characters of five different writers and seven different orientations (0, 10, 25, 30, 55, 75, and 90). The database prepare with these orientations for all five different writers. The fifty percent of the total database randomly selected used for training the classifier. The Table 8 shows the FMMN algorithm creates 270 hyperboxes and yields 100% recognition rate of training data set.

Table 8: Hyperboxes created and Recognition rates are obtained with the Training data set of moment of inertia based Radial Coding Features

Hyperbox size	No.of created	%Recognition
0.02	270	100

The total prepared database of hand written characters used for testing the FMMN algorithm. Table 9 shows FMMN algorithm yields 85% recognition rate for testing data set at which $\theta = 0.02$

Table 9: Hyperboxes created and Recognition rates are obtained with the Testing data set of moment of inertia based Radial Coding Features

Hyperbox	No. of created Hyperboxes	%Recognition rate
0.04	217	82.4
0.03	238	83.7
0.02	270	85.0

CONCLUSION

This work demonstrates a novel system to recognize the characters. The character recognition system is invariant to translation, rotation and scaling is reported to a very good recognition. This implemented recognition system based on FMMN classification algorithm is most robust and easy to implement. It is observed that the recall time of FMMN is small if the created hyperboxes are less. So choose the value of θ which will create less number of hyperboxes and recognize all the training patterns correctly.

REFERENCES:

[1] V.K.Govindan and A.P. Shivaprasad, "Character recognition –Review" pattern recognition, vol-23, no:2 pp 671-679 (1990).

[2] Jacek M. Zurada. "Introduction to artificial neural systems", jaico publishing house, 1999.

[3] H. Kauppinen, T. Seppänen, and M. Pietikäinen, "An experimental comparison of autoregressive and fourier- based descriptors in 2D shape classification," IEEE Trans. Pattern Anal. Machine Intell., vol. 17, pp.207– 210, Feb. 1995.

[4] G. L. Giles and T. Maxwell, "Learning, invariances, and generalization in high-order neural networks," Appl. Opt., vol. 26, no. 23, pp. 1972–4978, 1987.

[5]. C.-H. Teh and R. T. Chin, "On image analysis by the methods of moments,"IEEE Trans. Pattern Anal.Machine Intell., vol. 10, pp. 496– 513, July 1988.

[6] L.A.Torres-Mendez, J.C.Ruiz-Suaraz, Luis E.Sucar and G.Gomez, "Translation, Rotation and Scale Invariant Object Recognition", IEEE trans.on systems, man and Cybernetics, vol.30, No.1, February 2000.

[7] Rafel C.Gonzalez and Richard E.woods, "Digital Image Processing", Addison Wesley publishing company.

[8] P. K. Simpson, "Fuzzy min-max neural networks – Part 1: Classification," IEEE Trans. on Neural Networks, Vol. 3, No. 5, pp. 776-786, 1992.

[9] S.J Perantonis and P. J. G. Lisboa, "Translation, rotation and scale invariant pattern recognition by high-order neural networks and moment classifiers," IEEE Trans. Neural Networks, Vol. 4, pp. 276-283, July 1993

SIGNIFICANCE OF PROFESSIONAL COMMUNICATION AMONG ENGINEERING STUDENTS

Dumpala Nirmaladevi [1], V. Shrehari Anandh Reddy [2] , Nampelly Mallesham [3]

[1]* Assistant Professor of English, MALLA REDDY COLLEGE OF ENGINEERING FOR WOMEN, Hyderabad
Mail id: dumpalanirmaladevi@gmail.com
[2] V Shrehari Anandh Reddy, MALLA REDDY COLLEGE OF ENGINEERING, Hyderabad
Mail id: vsanandhreddy@yahoo.com
[3] Nampelly Mallesham, MALLA REDDY COLLEGE OF ENGINEERING, Hyderabad
Mail id: nampellymallesham@gmail.com

Abstract:

Communication skills are indispensible to students at all levels and in all walks of life throughout their professional and personal life. English Language communication has to be a dominant force. English Language communication is an inter linking process in an educational organization that pleases psychologically, unites the students socially, develops them functionally and establishes them professionally. It is an individual student necessity and educational institute imperative. Communication is a natural, ubiquitous behavioral process. Communication is the enabler of all management processes and functions. Generally students go on pep talk among themselves whiling away time or sharing pleasure or pain.

Sometimes students focus on baseless information created out of emotional thinking as sort of fanciful speculation. It is high time for the students to divert their unprofessional communication to professional communication. In view of the growing importance of English as a tool for a global communication and the consequent emphasis on training students to acquire language skills, the syllabus of English has been designed to develop linguistic and communicative competence of professional students. Communication is an important skill that not only reveals students' personality but also their domain knowledge, etiquette, and soft skills, etc. They develop professionalism among the students to communicate effectively with other peers. They need empathy, friendliness, professionalism in their speaking and attitude while interacting with others.

KEY WORDS: indispensible, dominant, pleases, imperative, fanciful

I. INTRODUCTION:

A large number of students enter the job market with an Engineering Degree every year. With limited number of opportunities, communication and language skills play a key role in increasing the employability of these candidates. Besides, the use of English has extended to all spheres of present day living. The primary objective of this paper is to prepare the professional students to use effectively and

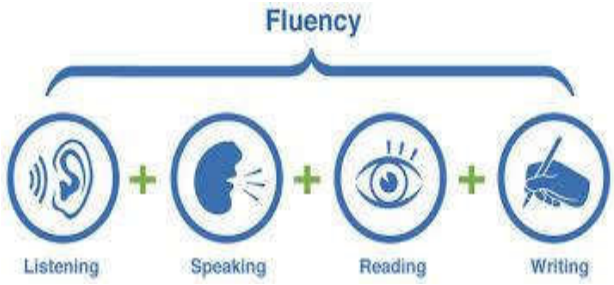
communicate confidently, especially in the context of engineering courses.

Objectives:

- Educate students in both the artistry and utility of the English language through the study of hands on learning through tasks, keeping theory based teaching to the minimum i.e. stressed more in the Text Book prescribed for engineering students throughout the country in the current academic year 2019-20.
- Students should be able to identify, analyze, interpret and describe the critical tasks, values, and themes that appear in literary and cultural texts and understand the way these ideas, values, and themes inform and impact culture and society, both now and in the past.
- Students should be able to ethically gather, understand, evaluate and synthesize information from a variety of written and various tasks in the Text Book, English for Engineers.
- Students must realize and adopt the policy of listening first anything next, then only they become good speakers. .

Communication skills are not only needed in day to day personal life, but also required at workplace and in professional life. We often find the certain situations which are challenging to handling, this is where having best communication skills become handy. Depending on the nature of our profession, we work in a team or interact with customers or other people.

"**I remind myself every morning: Nothing I say this day will teach me anything. So if I'm going to learn, I must do it by listening." – Larry King**



Every good communicator stresses on the importance of listening. With this basic lesson, let the professional students learn more to enhance their communication skills. Being successful at educational institution, workplace and in professional life with best communication skills make them progressive lifelong.

Let's get all the basics sorted first which provide skills for having best in communication, We can use these skills when we interact people, at workplace, in any profession and dealing with customers and others in professional life..

The one thing every great communicator does is start from scratch. Nobody is born great. We have to start from somewhere. It is best to start with basics. Understand what communication is and how we can effectively say what we intend to.

These are kinds of skills which are mentioned below. One great example is Steve Jobs' communication and negotiation style; he is regarded as someone who had the best presentation and negotiation skills in professional life. One of the great instances of having the best communication skills was the success of **Steve Jobs, Elon Musk** and others, although it's not the only factor which makes them successful, but definitely it's one of it.

11 ways to make the professional students competent in communication skills

1. Practice makes a student stand out of the crowd

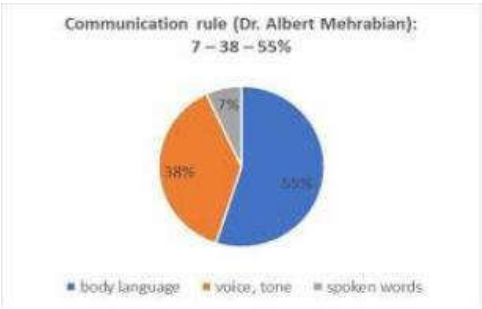
Pretty old but yet valid, this can take us a long way. The best approach is to practice it a lot. Whatever a student intends to do, let him rehearse it. While practicing, the faculty'll be able to assess students' performance.

2. Coordinate body and words

It is highly imperative that body says whatever we are saying. Body language including gestures and overall posture must be reflecting what we are trying to communicate. It is very unsatisfying if our body and words don't coordinate.

3. Maintain eye contact

Eye contact also has a great value. When we maintain eye contact with audience while communicating with them, it reassures that we are honest and sincere. Audience tends to believe us more when we maintain eye contact with them. Body language comprises more while speaking with any one on any occasion.



4. Speak slowly

It literally doesn't mean that we bore the audience with our low pitch and drooling speed. We just have to ensure that our speed is balanced and people can clearly listen each word we are saying.

5. Engage the listener

Engaging the listener means communicating in such a manner that people are interested in our expression. This is essentially important if we want to become an exceptional communicator.

6. Being a listener first

Listening helps us to understand what makes people listen. When we finally figure it out, we'll be able to make people listen to ours. Listening is the focus of every communication as it has a unique value in making us great communicators. Listening completely prepares us for answering in an effective way.

7. Tell stories (real ones)

Nothing gets the audience like stories and meaningful examples. We totally sell it when we tell a story. People listen to us more carefully and they become more interested. However, don't tell the fiction because when we tell real life stories and examples people understand that we are experienced and know what we are saying.

8. Let students participate

Be a friendly person who never gets offended when people ask questions. This not only makes the people interested but we get to learn and people value more. Nothing can prove our authenticity like open question and answer session.

9. Polish it

At this stage we finally give that pro touch that differentiates us from all mediocre conversations. Remember that we can't reach this stage without working on basics. Polishing is only required if we have certain skills.

10. Use humor

Humor is important if we want to sell our stories. Humor helps in impressing listeners. Humor validates our stories and polishes our existing skilsl. We should never worry to be a little clever and use humor. We'll shine more with humor.

11. Maintain positivity

This comes with great smile and exceptional attitude. People should feel comfortable and motivated when talking to us because we have this great attitude named positivity.

Strategic teaching techniques for effective communication skills for professional students.

- **Motivation:** Motivated students are an asset to an educational institution. Effective communication centers round the usage of words, speed of delivery of words, pitch modulation and body language. Using the right tools to communicate the right messages at the right time can salvage a crisis and motivate students to work towards success. Motivated students ensure not only the smooth and efficient functioning of project work but also increased congenial relationships with others.

- **Raising Morale** – Morale means "the capacity of a group of people to pull together persistently and consistently in pursuit of a common purpose". It is the product of motivation, which leads to zeal for action. The morale of the students fluctuates time to time. The faculty should therefore use effective communication to keep their morale high.

- **Inculcating the habit of Managerial efficiency**: - Communication helps in smooth operation of management. Managerial task can only be performed when communication system is effective.

- **Better decision**: – The success of any professional student can be measured in better decision. When the information, data and other facts are not effectively communicated, it hampers the decision making. So, when the facts are communicated to concerned peers, educational establishments and related persons. It is easy to make decisions promptly.

- **Removing controversies**- Effective communication allows smooth functioning that results conflict, controversies and disagreement can be resolved easily in students' professional as well as personal life.

Final word

There is always a room for improvement when it comes to communication skills. Thus make sure that we never stop taking these skills to the next level of improvement.

REFERENCE:

1. The Key Role of Communication Skills in the Life of Professionals by Rajendra Mahajan

2. English for Engineers by N P SUDHARSHANA and C. SAVITHA

(Cambridge University Press)

A REVIEW ON NANO PARTICLES, THEIR PHASES AND HEALTH EFFECTS.

1) **Grandhe.Radhika**
Assistant Professor
Malla Reddy College Of Engineering
grandhe.radhika@gmail.com

2)**K. Priya Dharsini**
Assistant Professor
Malla Reddy College Of Engineering
dharshinionline@gmail.com

3) **Dr.S.Nagaveni**
Associate Professor
MRITS,nagaveni.sangisetti@gmail.com

4) **Dr. T. Ram Prasad**
Assistant Professor
MRITS, talluriramprasad123@gmail.com

ABSTRACT:

This study observes that Nano material's are an active of research but also an economic sector which addresses many application domains in various fields. Nano materials are chemical substances that are engineered with particle sizes between 1 to 100 nanometers in at least one dimension. It can be established that engineered nano materials derive many functional advantages from their unique properties. Free nano particles are formed through various process like breaking down of larger particles or by controlled assembly processes. Man made nanoparticles engineered to have the different desired size, desired chemical composition, different surface nature in solid phase and charge properties can be produced in the liquid phase mainly through controlled chemical reactions. The main nature of formation of nanoparticles in the gas phase is by a chemical reaction in which gases are converted into tiny liquid droplets, followed by condensation and growth. Natural phenomena and many human industrial and domestic activities, manufacturing or road and air transport release certain amount of nanoparticles into the atmosphere, In advance Nano particles are used in Technological applications with magnetic storage, electro communications ,microwave devices and Medical applications such as drug delivery systems .This paper addresses in particular difficulties in defining these materials,the state of knowledge on human or environmental toxicity and agencies in change of safety.

KEYWORDS : Nan Materials, Nano Particles, Phases, toxicity, Applications.

I. INTRODUCTION:

Nanomaterial's and nano manufactured goods represent areas of scientifiic research and industrial application in full expansion. They are already an industrial and economic reality. As far as industrial sectors, the introduction of these products should be considered in terms of potential effects of toxicity and other material[1]. Free nanoparticles are formed through either the breaking down of larger particles or by controlled assembly processes.Natural phenomena and many human industrial and domestic activities, such as cooking,

manufacturing or road and air transport release nanoparticles into the atmosphere[3].
In recent years, nanoparticles intentionally engineered for advanced technologies and consumer products have become a new source of exposure. At present it is not clear just how significantly human exposure to these engineered nanoparticles has increased, be it in the workplace, or through the use of nanotechnology-based products [4].

II. DEFINITION OF NANOMATERIAL'S

"Nanotechnology - terminology and definition for the "Nano objects", i.e. Nanoparticle, nanofiber and nano plate", namely:

(i) nano- definition considered the range of dimension between 1 and 100 nm.

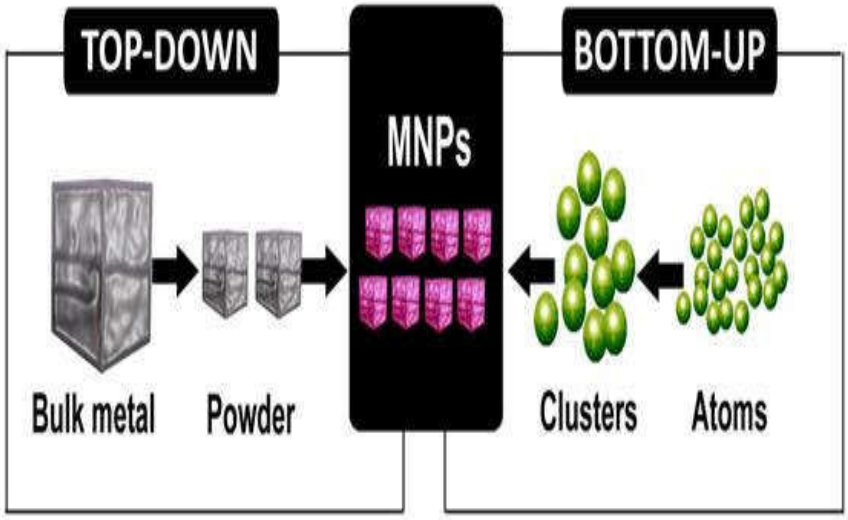
(ii) nano- objects are material with one, two or three external in the nano- domain dimension. Among these nano objects, nanoparticles include their three dimensions in the field of Nano; nano- oplates have one Nano sized dimension[2]

''Nanomaterial: means a material that meets at least one of the following criteria:

(i) Consists of particles ,with one or more external dimensions in the size range 1 nm- 100 nm for more than 1% of their number size distribution.

(ii) May have internal or surface structures in one or dimensions in the size range 1 nm-100 nm.

There are two approaches for the manufacturing of nanomaterials:



- The " **top-down** " approach involves the breaking down of large pieces of material to generate the required nanostructures from them. This method is particularly suitable for making interconnected and integrated structures such as in electronic circuitry.In the " **bottom-up** " approach,

single atoms and molecules are assembled into larger nanostructures. This is a very powerful method of creating identical structures with atomicprecision, although to date, the man-made materials generated in this way are still much simpler than nature's complex structures [5-7].

Sources of free nanoparticles:

Nanoparticles are formed through the natural or human mediated disintegration of larger structures or by controlled assembly processes. The associated processes occur either in the gas phase, in a plasma, in a vacuum phase or in the liquid phase, eventually followed by the intentional or unintentional transfer into one or more relevant fluid media and then to an individual receptor in an exposure setting [8-10].

Formation of nanoparticles in the liquid phase

Defined bottom-up production of nanoparticles in the liquid phase with respect to particle size, chemical composition, surface and charge properties occurs mainly through controlled chemical reactions (Frens 1973), and self limiting self assembly processes have evolved by controlling growth conditions. In view of the ecological cycling of nanomaterials, some emphasis has to be given to the corrosion and disintegration of bulk materials, where little knowledge is currently available (Oberdörster G et al 2005]. Naturally occurring processes generating nanosized structures in the liquid phase include erosion and chemical disintegration of organic (plant or microorganism debris) or geological (e.g. clays) parent materials. In all these types of disintegration process, the surface properties and their change through chemical reaction are critical in determining whether individual nanoparticles will be formed in the respective medium [11-14].

Formation of nanoparticles suspended in the gas phase

The main route of bottom-up formation of nanoparticles in the gas-phase is by a chemical reaction leading to a non-volatile product, which undergoes homogeneous nucleation followed by condensation and growth. Recently, this has become an important pathway for the industrial production of nanoparticle powders, which may be of metals, oxides, semiconductors, polymers and various forms of carbon, and which may be in the form of spheres, wires, needles, tubes, platelets or other shapes. This is also the unintentional pathway by which nanoparticles are formed following the oxidation of gas-phase precursors in the atmosphere, in volcanic plumes, in natural and man-made combustion processes, or in fumes associated with any man-made process involving volatilizable material at elevated temperature, such as welding or smelting, polymer fabrication, or even cooking .

As with the liquid phase case, disintegration processes of parent materials provide a pathway which only leads to nanoparticles suspended in the gas phase under special conditions. While in the liquid phase the presence of emulsifying agents accompanying an erosion or chemical disintegration process could support the suspension process, the dispersion of nanoparticles into a gas from liquid emulsions or dry powders is severely

limited by the strong adhesive forces between individual nanoparticles[15]. Therefore, any mechanically induced stress on the parent material mostly leads to particles in the micrometer range and above. Only under accidental conditions, e.g. in the case of uncontrolled release of a powder or an emulsion from a highly pressurized vessel could strong shear forces overcome these adhesive forces (Reeks and Hall 2001). In contrast, the spraying of liquids containing nanoparticles or soluble material at very low concentrations, followed by drying of the solvent, can lead to the resuspension of nanoparticles or to the formation of new nanoparticles from the solutes. This can lead to redistribution of nanoparticles, biological material or toxic substances into nanoparticulate airborne form[16].

The Sources of airborne nano particles:

In ambient air, the number of nanoparticles can be surprisingly similar in urban and rural areas, with as much as one million to one hundred million nanoparticles per litre of air depending on conditions.

In rural areas, nanoparticles are the product of chemical reactions involving compounds emitted by living organisms or by human activities such as wood burning.

In urban areas, most nanoparticles come from diesel engines or cars with defective or cold catalytic converters

In urban areas, the primary sources of nanoparticles are diesel engines or cars with defective or cold catalytic converters. Particularly, high-speed road traffic produces high numbers of nanoparticles of very small size.

In some workplaces, airborne nanoparticles may represent a potential health risk[17]. It is unlikely that nanoparticles would be released during manufacture because processes that generate them are often performed in closed chambers. Instead, exposure to nanoparticles is more likely to happen after the manufacturing process itself, or as a result of leaks arising from improper sealing. It is important to bear in mind that smaller nanoparticles remain airborne for longer periods of time than larger particles

III. OCCUPATIONAL SOURCES OF AIRBORNE NANOPARTICLES

Inhaled nanoparticles may represent a potential health risk. Aerosols in workplace environments may be derived from a wide variety of sources, depending on the type of activity and processes taking place. Nanoparticle aerosols arising from mechanical processes (e.g. the breaking or fracture of solid or liquid material) are unlikely to be formed. Grinding and surface finishing typically releases micrometre and submicrometre particles, possibly down to 100 nm but rarely below this. Most plasma and laser deposition and aerosol processes are performed in evacuated or at least closed reaction chambers. Therefore exposure to nanoparticles is more likely to happen after the manufacturing process itself, except in those cases of failures during the processing

(Luther 2004). In processes involving high pressure (e.g. supercritical fluid techniques), or with high energy mechanical forces, particle release could occur in the case of failure of sealing of the reactor or the mills. Nanoparticles exhibit increased diffusivity with decreasing size and therefore show delayed sedimentation in the earth's gravitational field, which translates into potentially increased lifetimes for nanoparticulate impurities at low concentration. In the presence of larger microparticles, as with the wide size distribution in aerosols such as smoke, the highly diffusive character of nanoparticles may lead to faster agglomeration or impaction on the larger particles. Furthermore, many particles, including metallic particles, are highly pyrophoric and there is a considerable risk of dust explosions.

IV. CHARACTERISTICS OF NANOPARTICLES ARE RELEVANT FOR HEALTH EFFECTS:

Toxicology of Nanoparticles:

Studies specifically dealing with the toxicity of nanoparticles have only appeared recently and, although now emerging in the literature, are still rare. Data concerning the behaviour and toxicity of particles mainly comes from studies on inhaled nanoparticles[18-20].

Health implications of nanoparticles used as drug carriers:

Particles for Drug Delivery

Carriers for Drug Delivery

Nanostructures and nanoparticles can be used for drug delivery purposes, either as the drug formulation itself or as the drug delivery carrier Current research focuses on cancer therapy, diagnostics and imaging, although many challenges still need to be solved .In addition, nanostructures are being investigated for gene delivery purposes Many different formulations involving nanoparticles have been used for drug delivery purposes, including albumin , poly(D,L-lactic-co-glycolide)acid (PLGA) (Panyam et al 2002, Weissenbock et al 2004), solid lipid formulations , cetyl alcohol/polysorbate nanoparticles ,hydrogels ,gold , polyalkylcyanoacrylate composites, magnetic iron oxide methoxy poly(ethylene glycol)/poly(ε-caprolactone) and gelatin . Albumin nanoparticles are already the subject of clinical studies for anticancer drug delivery purposes .

The use of nano particles as drug carriers may reduce the toxicity of the incorporated drug,although discrimination between the drug and the nano particle toxicity cannot always be made.The structure and properties of gold nanoparticles make them useful for a wide array of biological applications. Toxicity, however, has been observed at high concentrations using these systems. Goodman et al (2004) demonstrated that for 2 nm gold particles cationic particles were moderately toxic, whereas anionic particles were relatively non-toxic. Such very small sized gold nanoparticles were found to be non toxic when administered to mice for tumour therapy[21-23].

V. NANO SAFETY SUBSTITUTION

It means:

1) Replace toxic substance by less toxic substances,

2) Change the physical nature of the material,

3) Change the type of application: this approach relates to the replacement of an application in powder or liquid spray by an application in liquid phase,

4) Eliminate nanoparticles as soon as they are no longer necessary,

5) Favor forms which are non - dispersible in atmosphere including suspension in liquid or Use masters-mixtures.

VI. TECHNOLOGIES

This means implementing technical protection measures that are designed, as well as possible, to establish a barrier between workers and potentially dangerous substances or processes. Thus one effectively eliminates hazard exposure[24-25].

The following approach should be assessed

1) Use closed systems,

2) Use unbreakable containers or double containers for storage and transport,

3) Manufacture and use the substance on a form which limits its dispersion,

4) Capture pollutants at emitting source,

5) Air filtering prior to release,

6) Separate work premises and adapt ventilation of the premises

CONCLUSIONS

All of the country available studies, in vivo and in vitro; highlight the existence biological effects of nanomaterials in terms of functional, inflammation, modulations at cellular level for whole body. Still, little data are currently available, and it seems urgent to deepen existing awareness on the mechanisms involved in the dispersal of nanomaterials in the body. The risk assessment must not only take into account, but also their behavior and all of their life cycle.

Moreover, it is essential to know levels and situations of exposure, and therefore the condition of manufacture and composition of the products containing nanomaterials. In the absence of regulatory obligation, industrialists are very reluctant to communicate this information. In addition, epidemiological surveillance of workers exposed to nanomaterials should contribute decisively to the improvement of knowledge about their health effects possible medium and long term.

REFERRENCES:

1. Ratner M., Ratner D. "Nanotechnology: A Gentle Introduction to the Next Big Idea" 2003, Prentice Hall PTR 1 st edition.

2. Fulekar M.H. "Nanotechnology: Importance and Applications" 2010, I. K. International Pvt Ltd. 3. Karkare M. "Nanotechnology: Fundamentals and Applications" 2008, I. K. International Pvt Ltd.

4. Wilson M., Kannangara K., Smith G., Simmons M., Raguse B. "Nanotechnology: Basic Science and Emerging Technologies" 2002, CRC Press.

5. Hornyak G.L., Tibbals H.F., Dutta J., Moore J.J. "Introduction to Nanoscience and Nanotechnology" 2008, CRC Press.

6. Pinna N., Mato K. "Atomic Layer Deposition of Nanostructured Materials" 2012, John Wiley & Sons.

.7. Rieth M., Schommers W. "Handbook of Theoretical and Computational Nanotechnology" 2006, volume 1, American Scientific Publishers.

8. Schwarz J. A., Cristian I.C., Karol P. "Dekker Encyclopedia of Nanoscience and Nanotechnology" 2004, volume 3, CRC Press.

9. Kiyoshi N., Hosokawa M., Naito M., Yokoyama T. "Nanoparticle Technology Handbook" 2012, Elsevier.

10.http://en.wikipedia.org/wiki/Nanotechnology

11.Zhong W-H. "Nanoscience and Nanomaterials: Synthesis, Manufacturing and Industry Impacts" 2012, DEStech Publications.

12.Vajtai R. "Springer Handbook of Nanomaterials" 2013, Springer Science & Business Media. 13.Mody V. V. Mody, Rodney S., Singh A., Mody H.R. "Introduction to metallic nanoparticles" 2010, J Pharm Bioallied Sci. 2(4): 282–289

14.Zhang H-X, Siegert U., Liu R., Cai W-B. "Facile Fabrication of Ultrafine Copper Nanoparticles in Organic Solvent" 2009, Nanoscale Res Lett, 4:705–708

15.Wu N., Fu L., Su M., Mohammed A., Wong K.C., Dravid V.P. "Interaction of Fatty Acid Monolayers with Cobalt Nanoparticles" 2004, Nano Lett., 4(2): 383-386..

16. McNeil SE. Unique benefits of nanotechnology to drug delivery and diagnostics. Methods in molecular biology (Clifton, N.J.). 2011;697:3–8.

17. Maynard AD, Warheit DB, Philbert MA. The new toxicology of sophisticated materials: nanotoxicology and beyond. Toxicological Sciences. 2011 Mar 1;120 Suppl :S109–29.

18. Oberdörster G, Oberdörster E, Oberdörster J. Nanotoxicology: an emerging discipline evolving from studies of ultrafine particles. Environmental health perspectives. 2005;113(7):823–39.

19. Monteiro-Riviere NA, Tran CL. Nanotoxicology: characterization, dosing and health effects. informa healthcare.

20. Nanotoxicology: Laying a Firm Foundation for Sustainable Nanotechnologies, Nanotoxicology: Characterization, Dosing, and Health Effects, Informa Healthcare .

21. Isayev O, Rasulev B, Gorb L, Leszczynski J. Structure-toxicity relationships of nitroaromatic compounds. Molecular diversity. 2006;10(2):233–45.

22. Kane AB. Epidemiology and pathology of asbestos-related diseases. Reviews in Mineralogy and Geochemistry.1993;28(1):347–59.

23. Derfus AM, Chan WCW, Bhatia SN. Probing the Cytotoxicity of Semiconductor Quantum Dots. Nano Letters. 2004;4(1):11–8.

24. Donaldson K, Stone V, Tran CL, Kreyling W, Borm PJA. Nanotoxicology. Occupational and environmental medicine.2004;61(9):727–8. Introduction 26

25. Biotoxicity of metal oxide nanoparticles. Nanotechnologies for the Life Sciences. 2006.

SURFACE MODIFICATION OF BRASS FOIL

1. DEVARAKONDA SANDHYA RANI
Assistant Professor, H&S
Malla Reddy College of Engineering,
E-mail: ranisandhya131@gmail.com

2) YANDAMURI KHANJA LOCHANA DEVI
Assistant Professor, H&S
Malla Reddy College of Engineering,
E-mail: lochanadevi11@gmail.com

ABSTRACT

A copper-based alloy is which the main additive is zinc (up to 50 percent). Brass is an alloy of copper and zinc. It has good hot and cold pressure workability, excellent mechanical properties, attractive color, and relatively low cost [1]. Brasses are resistant to corrosion in many media. Brasses are particularly susceptible to corrosion by solutions containing ammonia or amines. Alloys with more than about 15% of zinc may suffer dezincification, which leaves a weak, porous corrosion deposit of copper. Stress corrosion cracking, particularly by ammonia and amines, is also a problem with the brasses. Alloys containing more than about 15% zinc are most susceptible. Use of the annealed temper, and annealing or stress relieving after forming, reduces susceptibility to stress corrosion cracking. We have different types of brass materials like alpha, alpha-beta, beta and white. Classification of brass is depending upon the composition of copper and zinc. In this project the structure of untreated brass foil and brominated brass foil is studied. Here we used different experimental techniques like Energy Dispersive X-ray Spectroscopy (EDS), x-ray Diffractometer and field emission scanning electron microscope. Elemental composition is known by EDS data, structure of foil is known by XRD data [3], FESEM produces clearer, less electrostatically distorted images.

KEY WORDS:

Brass foil, crystal structure, Energy Dispersive x-ray Spectroscopy (EDS), X-ray Diffractometer and Field Emission Scanning Electron Microscope.

I. INTRODUCTION:

Classification of brasses: different types of

Class	Copper (%)	Zinc (%)	Notes
Alpha brasses	>65	<35	Alpha brasses are malleable, can be worked cold, and are used in pressing, forging, or similar applications. They contain only one phase, with cubic crystal.
Alpha-beta brasses	**55–65**	**35–45**	**Also called duplex brasses. Suited for hot working. It contains both α and β' phase; the β'-phase is body-centered cubic and is harder and stronger than α. Alpha-beta brasses are usually worked hot.**
Beta brasses	50–55	45–50	Can only be worked hot, and are harder, stronger, and suitable for casting.
White brass	<50	>50	Too brittle for general use. The term may also refer to certain types of nickel silver alloys as well as Cu-Zn-Sn alloys with high proportions (typically 40 %+) of tin and/or zinc, as well as predominantly zinc casting alloys with copper additive.

The classification is clearly explained in this table.

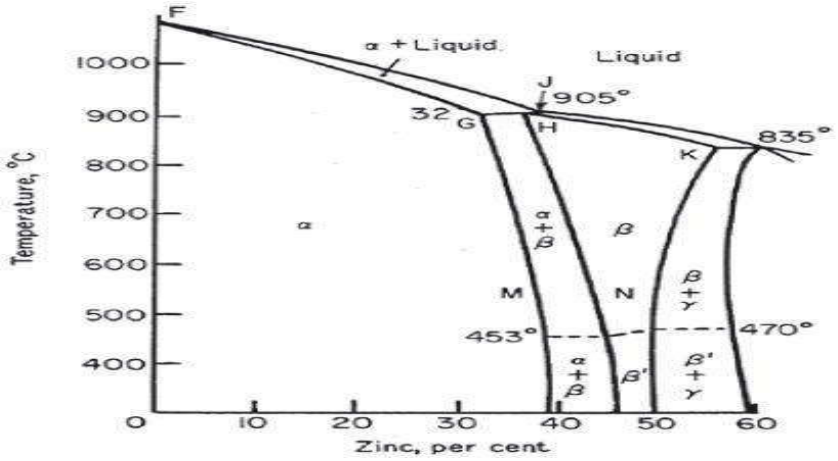


Fig. Phase diagram of Brass

The phase diagrams [5] of the different types of brasses are demonstrated in the following image.

In this phase diagram [5], α- phase is FCC,

β &β' are BCC, γ is a complex structure.

II. EXPERIMENTAL TECHNIQUES:

Various techniques were used to study the different properties of the material while doing the project. These techniques include X-ray Diffractometer (XRD) to determine the structural properties, field emission Scanning electron microscope (FESEM) and Energy Dispersive X-ray
Spectroscopy (EDS). In this chapter each of the above techniques is explained briefly.

2.1.Energy Dispersive X-ray Spectroscopy (EDS)

EDS identifies the elemental composition of materials imaged in a Scanning Electron Microscope (SEM) for all elements with an atomic number greater than boron. Most elements are detected at concentrations on the order of 0.1%.

2.2. Principle of Operation of SEM:

As the electron beam of the SEM is scanned across the sample surface, it generates X-ray fluorescence from the atoms in its path. The energy of each X-ray photon is characteristic of the element which produced it. The EDS microanalysis system collects the X-rays, sorts and plots them by energy, and automatically identifies and labels the elements responsible for the peaks in this energy distribution.

The EDS data are typically compared with either known or computer-generated standards to produce a full quantitative analysis showing the sample composition.

2.3. X-RAY DIFFRACTION:

The properties of a material can often be linked back to the arrangement of atoms in its crystal structure. X-ray diffraction [3] is a non-destructive analytical technique which can yield the unique fingerprint of Bragg reflections associated with a crystal structure.

One can regard a crystal structure as being built of layers, or planes, which each act as a semi-transparent mirror. X-rays with a wavelength similar to the distances between these planes can be reflected such that the angle of reflection is equal to the angle of incidence. We call this behavior `diffraction´ and it is described by Bragg's Law.

$$2d\sin\theta = n\lambda$$

When Bragg´s Law is satisfied, constructive interference of diffracted X-ray beams occur and a `Bragg reflection´ will be picked up by a detector scanning at this angle. The positions of these reflections tell us about the inter-layer spacing of atoms in the crystal structure, thanks to Bragg´s Law. Peak intensities give information about how much X-ray scattering is contributing to that reflection.

2.4. Field Emission Scanning Electron Microscopy (FESEM):

Principle of Operation:

A field-emission cathode in the electron gun of a scanning electron microscope provides narrower probing beams at low as well as high electron energy, resulting in both improved spatial resolution and minimized sample charging and damage. For applications which demand the highest magnification possible, we also offer In-lens FESEM.

Why Field Emission SEM?

- FESEM produces clearer, less electro statically distorted images with spatial resolution down to 1 1/2 nm. That's 3 to 6 times better than conventional SEM.
- Smaller-area contamination spots can be examined at electron accelerating voltages compatible with Energy Dispersive X-ray Spectroscopy.

Results and Discussions: EDS DATA

BRASS FOIL (NORMAL): TABLE 1

Element	Weight %	Atomic %
Cu K	60.66	61.34
Zn K	39.34	38.66
Total	100.0	

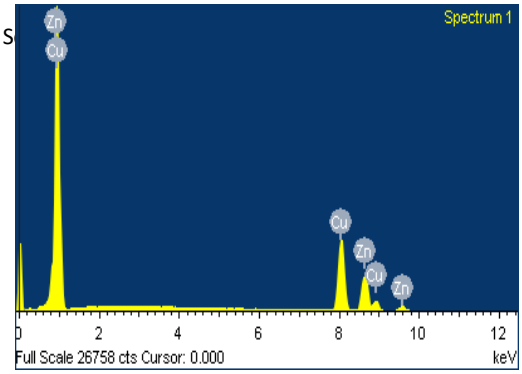


Fig1a:Shows the composition of cu and zn

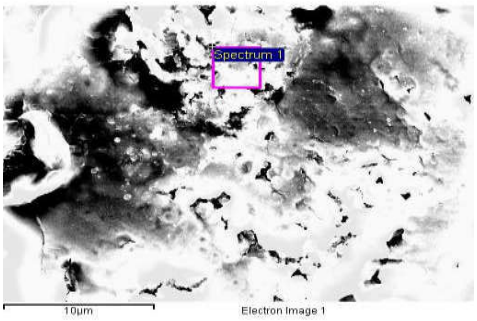


fig 1b: Shows the Morphology of Brassfoil

Element	Weight%	Atomic%
C K	3.12	12.67
O K	9.25	28.24
Ca K	0.71	0.86
e K	3.50	3.06
Cu K	4.68	3.60
Zn K	25.21	18.84
Br L	53.54	32.73
Totals	100.00	

Table 2: BROMINATED BRASS

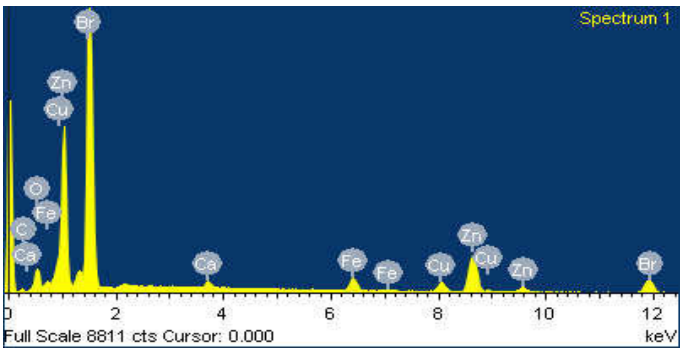


Fig 2a: shows the % of elements after bromination

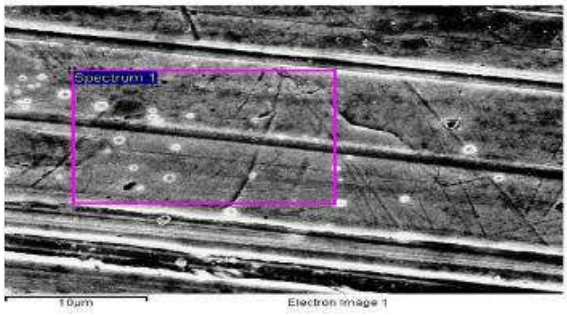


Fig: 2b Morphology of brominated brass

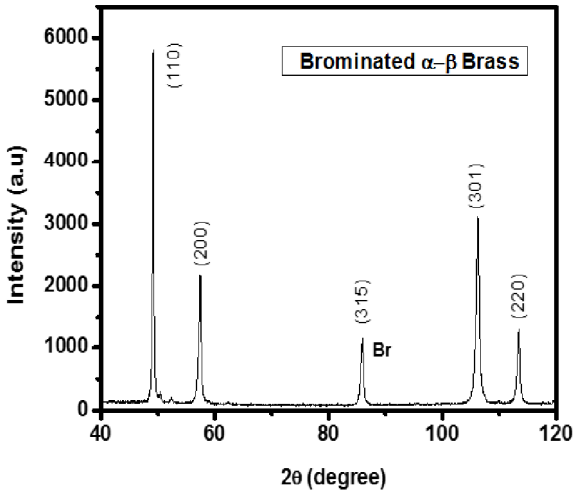


Fig 3b. XRDpattern of brominated α-β brass

Crystallite sizes and lattice parameters [4] found as 36 nm.

III. XRD data:[3] PURE BRASS & BROMINATED BRASS

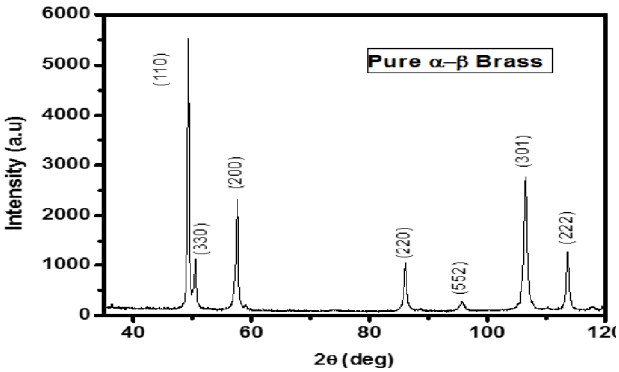


Fig 3a. XRD pattern of pure α-β brass

Clearly shows the α-β brass with cubic structure Crystallite sizes and lattice parameters [4] found as 27 nm.

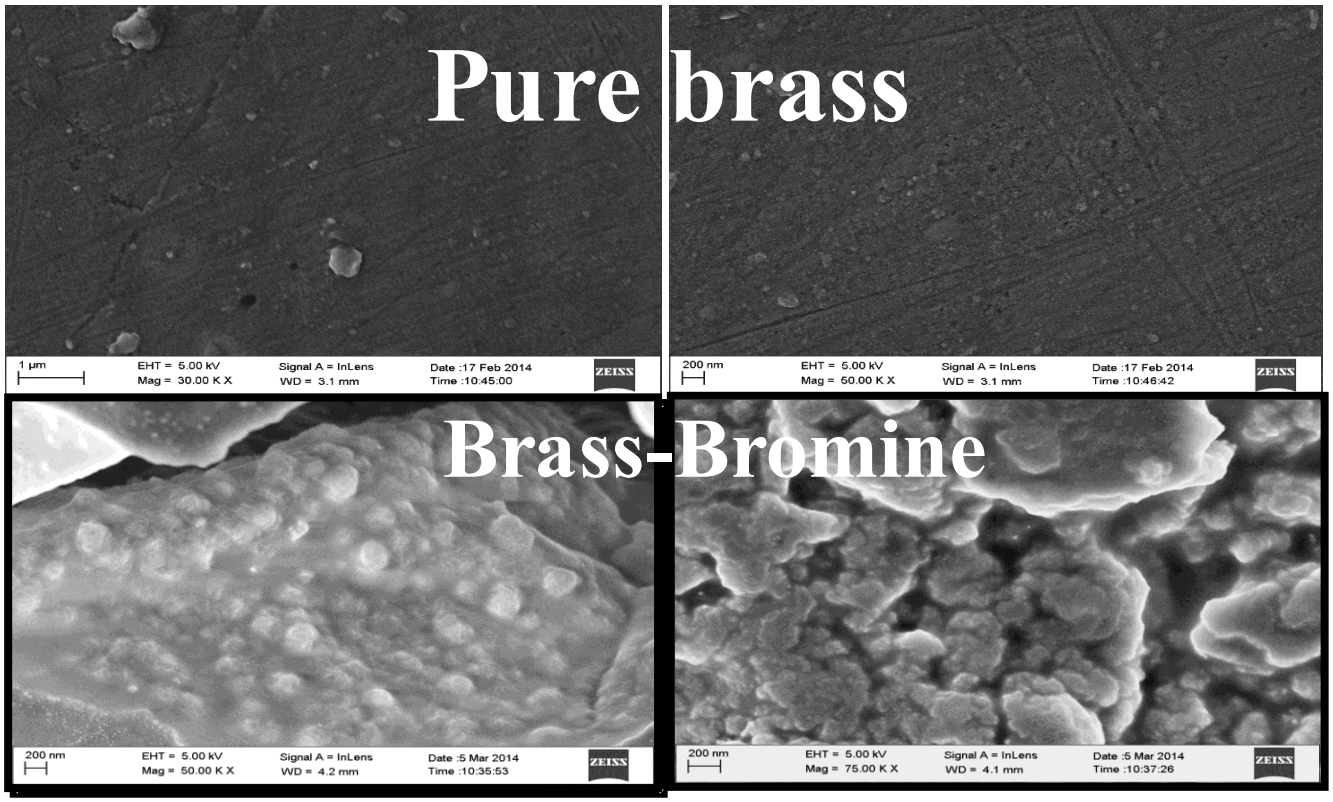


Fig 4. FESEM images clearly show s pure brass foil and brominated brass foil.

FESEM images of pure brass shows a microstructure with highly visible cracks in low and high magnification also but in case of bromine treatment we observed small cluster nanoparticles with 50-80 nm regimes. In figure Bromine brass shows very good image with clean morphology of nano-flake type morphology [6] with controlled dimension.

IV. CONCLUSIONS

The objective of the work in this project to modify the surface of commercial brass foil. Initial thickness of the brass foil is 0.4mm.To do this modification bromination technique was used. Initial structure of this foil without any treatment is cubic, after bromination structure changes to hexagonal.

From EDS data I deduced the composition of copper and zinc in the foil. From XRD, I knew the structure and crystallite size of my foil before and after treatment. From FESEM, I knew the morphology of my sample.

This preliminary work provides motivation for a detailed future study of structure, phase transition and phase stability of surface modified brass.

REFERENCE:

[1]. Smiriagin, A. P. Promyshlennye tsvetnye metally I splavy, 2nd ed. Moscow, 1956.

[2]. Khurmi, R.S. and Gupta, J.K. (2004) engineering designer, 30, 6 and J.K.Gupta 44120476

[3]. Swanson, H.E., Gilfrich, T. and Ugrinie, M. (1955) Standard X-Ray Diffraction Powder patterns.

[4]. Crystallography for solid state physics Lattice Geometry: Crystallographic formulae from appendix 3.2 pg.no.94.

[5]. Substitutional solid solutions –Home Ruthery rules –chapter 22, pg.no.625. [Phase diagram]

[6]. Crystallography applied to solid state physics by A.R.Verma & O.N. Srivastava from chapter 3, pg.no. 65 [morphology and angular relationships].

ASSESSMENT OF INDUSTRIAL EFFLUENTS AND ITS IMPACT ON GROUND WATER QUALITY IN AND AROUND BALANAGAR, HYDERABAD,A.P, INDIA

1.Sandhya Rani S, **2. Madhavi A**

1.Assistant Professor, Department of chemistry, Mallareddy College Of Engineering
E-mail: [1]sandhyasaro9@gmail.com

2. Assistant Professor, Department of chemistry, Mallareddy College Of Engineering
E-mail: [2] madhavi.avirineni@gmail.com

Abstract:
Due to rapid industrialization and overexploitation of the groundwater resources, there is a drastic change taking place in the Hyderabad urban environment. The present study area is Jeedimetla of Hyderabad, AP, India. The possible quality of the heavy metals major and trace elements in the hydrological system have been identified and quantified in the present study.

Depending on the environmental parameters, the trace metals and associated elements may form complexes and precipitate or become concentrated at several places. In the present study is an attempt made to evaluate the changes in the water quality of the Jeedimetla area.All trace elements, which include As, Se, B, V, Fe, Co, Pb, Cu, Zn, Cd, Mn, Ni, Mo, Ba, and Na, Mg, Si, Ca, Al, were analyzed using the highly sensitive inductively coupled plasma-mass spectrometer (ICP-MS). For most of the elements, the detection limits are around 1ppb.

Analytical data were processed with various computer programs for statistical evaluation. The impacts of natural and anthropogenic sources on the elemental concentration and the total area affected by each element have been deciphered using these graphs. The range and average concentrations of some of the trace elements of environmental concern in the Ground water.

The average concentrations are listed, and the world health organization (WHO) and minimum.National Standards (MINAS), Central pollution control board (CPCB) for water quality are also listed for comparison. Most of the trace elements show high concentrations (several orders of magnitude) in the groundwater of Jeedimetla in comparison to global average standards. Especially salts like Na, Ca, Mg, Se, are at peak levels and add turbidity to water and make it worst.

Key Words: environmental pollution, anthropogenic, irrigation practices, hydrogeochemistry

1. INTRODUCTION

Water is very vital for nature and can be limiting resource to men and other living beings. Without a well functioning water supply, it is difficult to imagine productive human activity be it agriculture or livestock. Water quality is influenced by natural and artificial effects including local industrial waste pollutants, geology and irrigation practices the hydrogeochemistry of water is important to quality in water supply planning for living areas. The geochemical character of any groundwater determines its quality and utilization. The quality is function of the physical, chemical and biological parameters

and it should be subjective, since which depends on a particular intended use.

The various on hydrogeochemistry of water quality have been carried out by various members. Laluraj et al.(2005) have been studied ground water chemistry of shallow aquifers in the costal zones of Cochin and concluded the ground water present in the shallow aquifers of some of the stations were poor in quality and beyond the potable limit as per the standard by World Health Organization (2004). Rapid increase of urbanization and industrialization leads to deterioration in hydrogeochemistry of ground water quality. Srinivas et al. (2000) and Jha and Verma (2000) have reported the degradation of the water quality in Hyderabad and Bihar, Respectively, Patnaik et al (2002) have studied water pollution generated from major industries similarly, waste pollutants or effluents. Discharged into streams may enter the aquifer body downstream. This also affects the ground water geochemistry. The studies on trace metals have been carried out (Jangir et al. 1990; Sharma et al.)Sharma et al (2004) Singh and Chandel (2003, 2006) pollution problems in ground water and industrial waste water have been studied. The specific objectives of the present study area 1) the investigations and interpretation of hydrogeochemistry of Jeedimetla and adjoining areas. 2) Find out the suitability of groundwater for irrigation and drinking purpose and 3) establish significant correlation among ton parameters of ground water samples.

Method

Balanagar area is located at Hyderabad city, in the NE of Andhra Pradesh state, is undergoing rapid urbanization and industrialization.

Groundwater samples collected from different hand pump, tube wells and ponds at 5 sampling points were analyzed. Samples were collected in good quality polythene bottles of 1-1 capacity. Sampling was carried out with out adding any preservatives in rinsed bottles for avoiding any contamination and brought to the laboratory. Only high pure (anal R Grade) chemicals double distilled water was used for preparing solutions for chemical analysis. Physical parameters like Ph, total dissolved solids, and electrical conductivity were determined at site with the help of digital portable analyzer kit. The samples collected and as per procedure (APHA 1995).

The total hardness (TH) in mg/l was determined by following equation (Todd 1980)

$$TH = 2.497\ Ca^{2+} + 4.115\ Mg^{2+}$$

Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
48.68	42.75	67.88	87.91	95.73

All the results compared with standard limits recommended by world health organization (WHO 2004).

2. AIMS AND OBJECTIVES:

- The present study area was undertaken with an objective to assess the quality of ground

water collected from Balanagar,Hyderabad,AP.

- To assess the quality of drinking water in the Balanagar .
- To determine various physic-chemical parameters in the water samples.
- To assess heavy metal concentrations in ground water and to locate the possible source of contamination.
- To study the impact of heavy metal pollutants on ground water and to know the health effects on the people living and using the water for different purpose.

3. MATERIALS AND METHODS:

METHODOLOGY

Sample collection

The total number of 5 sites were sampled during post-monsoon season (Feb 2013). Most of the water samples were collected from bore wells and acidified immediately to bring the pH of the solution to < 2.0. Clean polythene bottles of litre capacity soaked with 1:1 HNO3 and washed using detergent was used for groundwater sampling. These bottles were allowed to stand for several hours in double distilled water before taking to the field. The sample bottles were rinsed two to three times using the representative groundwater samples. Water samples were collected 30 cm below the water level in open wells using water sampler. Bore well water samples were collected directly from the pump sets. Since industrial pollutants contaminate the upper layer of soil, the samples were collected from wells that are being constantly used.

3. Sample preparation

All the samples were filtered in Whattman 42 filter paper and were diluted to 20 times for further analysis by ICP-MS. Care was taken to avoid contaminants to enter into the sample solution.

4. Sample analysis

Physical parameters of groundwater such as, pH and EC were determined in the field using digital meters immediately after sampling. Collected groundwater samples were transported to the laboratory on the same day. They were filtered using 0.45 microns Millipore filter paper and acidified wit[Abstract]h nitric acid (Ultrapore, Merck) for cations. For All trace elements (Li, Be, B, Si, V, Cr, Mn, Fe, Co, Cu, Zn, As, Se, Rb, Sr, Mo, Ag, Cd, Sb, Ba and Pb) were determined quantitatively by ICP-MS following Balaram (2003).

5. Instrumentation

A Perkin Elmer SCIEX, Model ELAN DRC II ICP-Mass Spectrometer (Concord, Ontario, Can-ada) was used throughout. The sample introduction system consisted of a standard Meinhard equalizer with a cyclonic spray chamber. All quantitative measurements were performed using instrument software. The software uses knowledge-driven routines in combination with nu-merical calculations to perform an automated interpretation of the whole spectrum. Several well-known isobaric interferences are programmed, and the corrections are automatically ap-plied.

Instrumental and data acquisition parameters are listed elsewhere (Balaram and Rao, 2003).

6. Calibration Strategies

In order to overcome matrix effects that are generally observed during the analysis of water samples, several methods can be used. Dilutions of the sample to bring down the total dissolved solids content to < 200 mg/L, use of internal standard for calibration etc., are effective for this purpose. By using alternative isotopes interferences can often be avoided, except for the mono-isotopic elements. NIST 1640 (certified reference material trace elements in natural water ob-tained from National Institute for Science and Technology, USA) was used to calibrate the sys-tem. The isotopes of measured elements in this work are free from potential polyatomic, iso-baric, and doubly charged ion interferences.

7. Sampling Area

The ground water samples from pumps were collected from the Balanagar, Hyderabad, Ap.

A total of 5 water samples were collected in presterilized bottles (schott duran et al. Germany)

And were stored at 2 to 4 degrees. The physicochemical properties such as hydrogen ion

Concentration (pH), total dissolved solids (TDS) in water sample were analysed on pH Bench top

Meter (Thermo electron corp. orion 5 star), using standard procedures. ICP-MS (perkin-Elmer

Sciex Elan drc II). ICP-MS is a type of producing ions (ionizations) with a mass spectrometer as a

Method of separating and detecting ions.ICP-MS is highly sensitive and capable of determination

Of a range of metals and several non-metals at concentrations

SAMPLING AREAS-TABLE 1.

S.NO	SAMPLE CODE	LOCATION AREA
1	SAMPLE 1	CITD, BALANAGAR
2	SAMPLE 2	NRSC, BALANAGAR
3	SAMPLE 3	IDS SCHOOL OF AUDITORIUM, BALANAGAR
4	SAMPLE 4	HAL AUDITORIUM, BALANAGAR
5	SAMPLE 5	BALAJI STEEL, BALANAGAR


Government of India
Department of Space
ISRO
NATIONAL REMOTE SENSING CENTRE
nrsc


हिंदुस्तान एरोनाटिक्स लिमिटेड
HAL
HINDUSTAN AERONAUTICS LIMITED

1. Tables and graphs

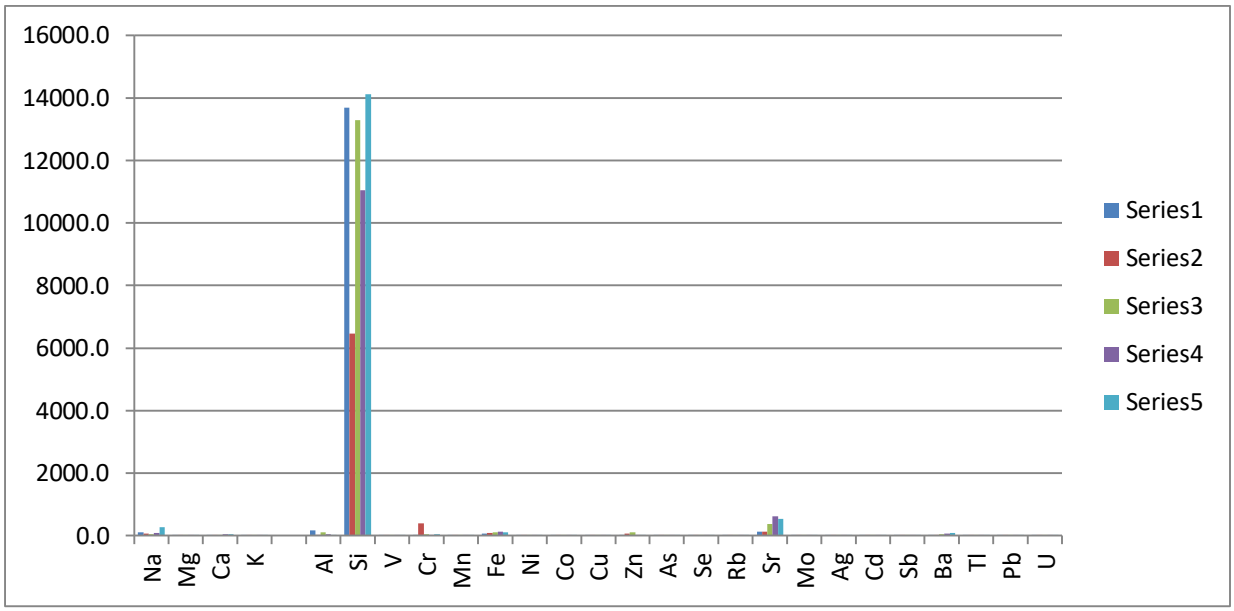
Table 1: Analytical data of major, minor and trace elements in groundwater samples collected from Balanagar, Ranga Reddy District, Andhra Pradesh.

ELEMENTS	unit	VA1	VB2	VC3	VD4	VE5
Na	ppm	111.3	52.69	41.85	83.82	257.6
Mg	ppm	6.241	5.766	15.66	25.51	24.06
Ca	ppm	10.35	11.11	24.88	33.09	35.71
K	ppm	1.625	1.341	1.400	1.286	1.591
Al	ppb	159.1	24.30	109.8	39.69	12.76
Si	ppb	13684	6454	13275	11044	14113
V	ppb	12.70	8.090	6.630	5.015	9.730
Cr	ppb	11.58	396.1	44.15	15.83	43.78
Mn	ppb	4.860	3.160	4.100	18.48	8.795
Fe	ppb	69.16	74.20	102.9	119.6	112.9
Ni	ppb	1.310	1.220	2.800	3.345	3.255
Co	ppb	0.175	0.095	0.265	0.425	0.775
Cu	ppb	6.770	19.51	12.75	7.745	19.92
Zn	ppb	11.63	65.77	104.5	17.76	25.28
As	ppb	1.420	1.255	0.710	1.150	2.695
Se	ppb	8.005	6.980	6.490	8.895	9.900
Rb	ppb	2.005	1.290	0.815	0.605	1.630
Sr	ppb	124.1	118.0	371.5	606.8	525.0
Mo	ppb	6.325	3.180	1.435	1.930	8.615
Ag	ppb	0.055	0.020	0.020	0.030	0.020
Cd	ppb	0.020	0.075	0.140	0.020	0.050
Sb	ppb	0.035	0.040	0.025	0.020	0.035
Ba	ppb	30.27	21.45	43.18	64.42	88.26
Tl	ppb	0.025	0.015	0.015	0.015	0.015
Pb	ppb	0.045	2.330	2.750	0.130	0.100
U	ppb	5.420	1.650	3.320	2.420	6.490

Table 2: Summary of the analytical data of major, minor and trace elements in groundwater samples collected from Balanagar, Ranga Reddy District, Andhra Pradesh

ELEMENTS	unit	min	max	mean	std	WHO	BIS
Na	ppm	41.85	257.6	109.4	87.17	<200	
Mg	ppm	5.766	25.51	15.45	9.407	<30	30
Ca	ppm	10.35	35.71	23.03	11.92	75	75
K	ppm	1.286	1.625	1.449	0.152		
Al	ppb	12.76	159.1	69.12	62.84	200	30
Si	ppb	6454	14113	11714	3170		
V	ppb	5.015	12.70	8.433	2.956		
Cr	ppb	11.58	396.1	102.3	164.9	50	50
Mn	ppb	3.160	18.48	7.879	6.302	500	100
Fe	ppb	69.16	119.6	95.74	22.82		300
Ni	ppb	1.220	3.345	2.386	1.044	20	
Co	ppb	0.095	0.775	0.347	0.269		
Cu	ppb	6.770	19.92	13.34	6.247	2000	50
Zn	ppb	11.63	104.5	44.98	39.41	3000	5000
As	ppb	0.710	2.695	1.446	0.746	10	50
Se	ppb	6.490	9.900	8.054	1.389		
Rb	ppb	0.605	2.005	1.269	0.574		
Sr	ppb	118.0	606.8	349.1	224.6		
Mo	ppb	1.435	8.615	4.297	3.074	70	
Ag	ppb	0.020	0.055	0.029	0.015		
Cd	ppb	0.020	0.140	0.061	0.050	3	10
Sb	ppb	0.020	0.040	0.031	0.008	50	
Ba	ppb	21.45	88.26	49.51	27.02	700	
Tl	ppb	0.015	0.025	0.017	0.004		
Pb	ppb	0.045	2.750	1.071	1.350	10	50
U	ppb	1.650	6.490	3.860	2.037	2	

GRAPH:



5. RESULTS AND DISCUSSION

Al:

ALLUMINIUM.In study area varys from 12.76 to 159.9with an average of 69.12.The permissible

limit of Al WHO 200 OR BIS 30. All the samples are exceeding the permissible limit except 2&5.

Si:

SILICON.In study area varys from 6454 to 14113 with an average Of 11714.The permissible limit of

Si WHO 0.05 .All the samples are exceeding the permissible limit. High Si in water Silicosis disease.

V:

VENEDIUM.In study area varys from 5.01 to 12.7with an average of8.43.The permissible limit

of V WHO 10. All the samples are below the permissible limit except sample1.

Cr:

CHROMIUM.In study area varys from 11.58 to 396.1 with an average of102.3.The permissible limit

of Cr WHO 50 OR BIS 50. All the samples are below the permissible limit except sample2.

Mn:

MANGANESE.In study area varys from 3.16 to 18.48with an average of 7.87.The permissible

limit of Mn WHO 500 OR BIS 100. All the samples are below the permissible limit.

Fe:

FERROUS.In study area varys from 69.16 to119.6 with an average of 95.74.The permissible limit of

Fe WHO 1-3 OR BIS 300. All the samples are below the permissible limit. High Fe in water cause siderosis.

Ni:

NIKEL.In study area varys from1.22 to 3.34 with an average of 2.38.The permissible limit of Ni

WHO 20. All the samples are below the permissible limit.

Co:

COBALT:In study area varys from0.09to 0.77 with an average of 0.34.The permissible limit of

Co WHO 0.2. All the samples are below the permissible limit except sample 4. . High Co in water

cause lazyness, headache, death.

Cu:

CUPPER:In study area varys from 6.77. to 19.92. with an average of 13.34.The permissible limit of

Cu WHO 1.5. All the samples are exceeding the permissible limit.

Zn:

ZINC:In study area varys from 11.63. to 104.5 with an average of 44.98.The permissible limit of Zn

WHO 500 OR BIS. All the samples are below the permissible limit.

As:

ARSENIC:In study area varys from0.71 to 2.69 with an average of 1.44.The permissible limit of

As WHO 10 OR BIS 50. All the samples are below the permissible limit. High As in water causes

cancer.

Se:

SELENIUM:In study area varys from 6.49. to 9.90 with an average of 8.05.The permissible limit

of Se WHO OR BIS.

Rb:

RUBIDIUM:In study area varys from0.60 to 2.005 with an average of 1.26.The permissible limit of

Rb WHO OR BIS.

Sr:

STRANICIUM:In study area varys from 118.0to 606.8 with an average of 349.1The permissible

limit of Sr WHO OR BIS.

Mo:

MOLYBDENUM:In study area varys from 1.43to 8.61 with an average of 4.29.The permissible

limit of Mo WHO OR BIS.

Ag:

SILVER:In study area varys from 0.02 to 0.055 with an average of 0.02.The permissible limit of

Ag WHO OR BIS 0.2. All the samples are below the permissible limit except sample 1&4 .

Cd:

CADMIUM:In study area varys from 0.02to 0.01 with an average of 0.06.The permissible limit of

Cd WHO 3 OR BIS 10. All the samples are below the permissible limit. High Cd in water Itai Itai

,anaemia diseases.

Sb:

ANTIMONY:In study area varys from 0.02 to 0.04 with an average of 0.03The permissible limit

of Sb WHO 50 OR BIS -.All the samples are below the permissible limit.

Ba:

BARIUM:In study area varys from 21.45 to 88.26 with an average of 49.51.The permissible limit of

Ba WHO 700 OR BIS-. All the samples are below the permissible limit.

Pb:

LEAD:In study area varys from0.45to 2.75 with an average of 1.07.The permissible limit of Pb

WHO 10 OR BIS 50. All the samples are below the permissible limit.

Tl:

TALIUM:In study area varys from 0.01to 0.02 with an average of 0.01.The permissible limit of

Tl WHO OR BIS.

U:

URENIUM:In study area varys from 1.65 to 6.49 with an average of 3.86.The permissible limit of

U WHO 2 OR BIS 5. All the samples are below the permissible limit except sample1&5

5. CONCLUSIONS

The groundwater in the area is generally alkaline in nature and the pH varies from 7.2 to 8.61. Based on the concentration of TDS, 11 samles are with in the *"permissible limits"* both for drinking and irrigation while the remaining 3 samples useful only for *"irrigation"*. Based on Wilcox's (1948, 1955) classification 4 samples fall in the fields of excellent to good, 9 samples that in good to permissible and one sample falls out side of the figure indicates that it is not useful for irrigation.

Classification based on SAR alone for irrigation suitability indicates that all samples are suitable for irrigation purpose.

According to USSL classification, most of the samples belong to C_2S_1 and C_3S_1 category suggesting that the Rangapur water is suitable for agriculture. The Gibbs diagram indicates that majority of the groundwater fall in the rock dominant field.

The quality of groundwater is controlled by lithology apart from other factors like land use pattern. Based on the above observations it is noticed that the groundwater in the present study area except few samples are suitable for both drinking and agriculture. But it is observed that the people are suffering with endemic diseases and agricultural crop production is declined drastically. Detailed study of trace elements is expected to unravel the suitability of the water of various purposes.

BIBILOGRAPHY

1. *Balaram, V. 2003. Groundwater Analysis by coupled plasma-Atomic Emission Spectrometry (ICP-AES). Proceeding of workshop on Groundwater Contamination, Central Groundwater Board, New Delhi: 16-30*
2. *Balaram. V. & Gnaneshwar Rao. T. 2003. Rapid Determination of REEs and other Trace Elements in Geological samples by Microwave Acid Digestion and ICP-MS. Atomic Spectroscopy 24(6):206-212*

EFFECTIVE COMMUNICATION IN MARKETING

1) **G Avinash**
Student, Department MBA,
Malla Reddy College of Engineering,
Hydeabad.
Mail id: yabhi010@gmail.com

2) **N Manjari**
Student, Department MBA,
Malla Reddy College of Engineering,
Hydeabad.
Mail id: manjarinesha1997@gmail.com

ABSTRACT: Marketing Communication emerges as a powerful tool that guides practitioners in developing and implementing marketing communications more consistently and effectively. Despite its continuing appeal little is known about its physical or visible form in marketing communication process, but the emergence of this concept has become one of the most significant examples of development in the marketing discipline. It is the most innovative function of marketing endorsed by advertising and marketing practitioners. Marketing Communication has moved beyond communication to the process of using promotional elements in a unified way so that a synergistic communication effect is created and achieved. The present study explore on the purpose of drawing out whether Integrated Marketing Communication could indeed be seen as an effort towards promotion in practice and also find out some challenges and future of integrated marketing communications for business.

I.INTRODUCTION:

Integration occurs when separate people or things are brought together, like the integration of students from all of the district's elementary schools at the new middle school, or the integration of snowboarding on all ski slopes. You may know the word differentiate, meaning "set apart." Integrate is its opposite. Marketing communication includes Advertising, Sales Promotion, Events and Experiences (sponsorship), Public Relations and Publicity, Direct Marketing, Interactive Marketing, Word-of-Mouth Marketing, Personal Selling. These tools of communication are collectively called as Marketing Communication Mix, concept of integrated marketing communications involves the solution of two interrelated problems: creating a system of communication messages using different communications tools that would not contradict each other, which would be coordinated among themselves, forming a single favorable image of the communicator; maximize the effectiveness of communications tools by searching for optimal various means. We made the conclusion that the use of an integrated approach reduces the costs of the company, since the integrated marketing communications carefully coordinates and mutually links the use of all elements of the marketing mix.

II. KEY WORDS:

Integrated Approach, Marketing Mix, Communications Tools, Integrated Marketing Communication, System of Communication, Communicator

III. MARKETING COMMUNICATIONS TOOLS:

Each marketing communications tool has certain advantages along with disadvantages, each tool has a specific purpose in terms of achieving the goals of product promotion, and therefore, the most effective way of promotion will be integrated marketing communication. It is integrated marketing communications that allows manufacturers to coordinate and strengthen their communications in order to achieve maximum impact Combining various marketing communications tools (MK) allows you to get a synergy effect when the combined use of individual marketing tools (MI) allows you to achieve a greater result than with their separate use. The synergy effect is achieved by combining the advantages inherent in each communications tools separately. Integrated marketing communications rebuilds marketing communications in order to see them as they are seen by the consumer - as the flow of information from a single source. Thus, the concept of integrated marketing communications involves the solution of two interrelated problems:

- Creating a system of communication messages using different communications tools that would not contradict each other, which would be coordinated among themselves, forming a single favorable image of the communicator;
- Maximize the effectiveness of communications tools by searching for optimal various means. The development of an integrated marketing communications strategy consists of nine steps:
- Assessment of the situation: identification of possible difficulties and opportunities. In assessing the situation, the main attention should be paid to factors affecting the effectiveness of marketing messages. Marketing communications can solve the problems of image, consumer attitudes, product perception or dissemination of information about it.
- Goal setting: communication goals can be planned with a focus on the impact of the strategy being developed on the consumer. The objectives of marketing communications usually consist of several categories:
  A) Creating awareness;
  B) Achieving understanding
  C) Ensuring changes in the relationship eniya to the product and its perception;
  D) achieving changes in consumer behavior
  D) Reinforcements of previous decisions.
- Target audience selection: within integrated marketing communications, target audiences are identified for specific communication activities. At the same time, the target audience is much wider than the target market. For example, the target toy market is mainly children. At the same time, their parents, grandmothers, etc. may also be included in the composition of their target audience
- The choice of marketing communications mix. Different kinds activities used to achieve marketing communication goals form a communication mix. Composition communications mix is selected individually for different market

segments and different market situations. Communication tools may include advertising media, sales promotion, PR, direct marketing, personal selling, etc.

IV. UNDERSTANDING THE MARKETING – COMMUNICATION:

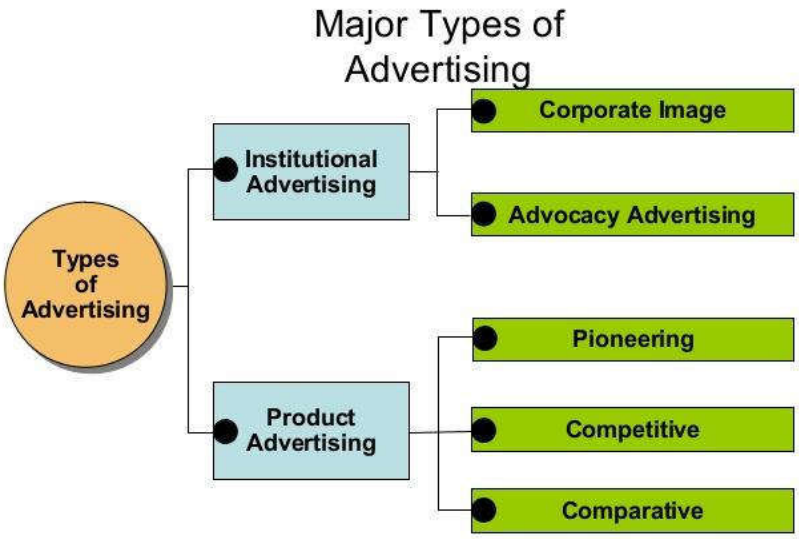
Positive effect on strengthening customer commitment. Recently there has been a tendency to reduce brand loyalty. This is due to the fact that the number of available goods is growing and at the same time sales promotion measures are activated. The unrestricted use of sales promotion measures ensures that even the most loyal consumers are beginning to expect special offers and lower prices from the company. Such excessive sales promotion creates for I companies have great difficulty in ensuring the commitment of customers to their brand name. In this case, the price factor becomes more important than the name factor of the manufacturer of the goods. Therefore, many companies use more complex marketing programs to strengthen consumer commitment to a particular brand than just advertising, such as direct marketing, public relations, organization of special marketing events, etc. New contacts are needed to implement bilateral contacts between the company and its target audience. Moreover, the initiative to establish contact can or should come directly from the consumer. In the near future, these types of communications will complement or even replace those types of mass communications that have been used in the recent past or are being used now. In this direction, many firms began to pay great attention to the use of the so-called relationship marketing relationship with the client). The use of relationship marketing allows you to identify current and potential customers of increased value and link them to the brand by providing each of them with special signs of attention.

Most company executives target marketing communications to meet the personal needs and desires of customers. Relationship marketing fits in well with integrated marketing communications, since a program to create strong customer relationships requires more than just advertising a product through the media. Its implementation requires the use of a fully integrated communication process. This should take into account the features of all types of marketing appeals and all aspects of the transfer of information about the company and its brand.

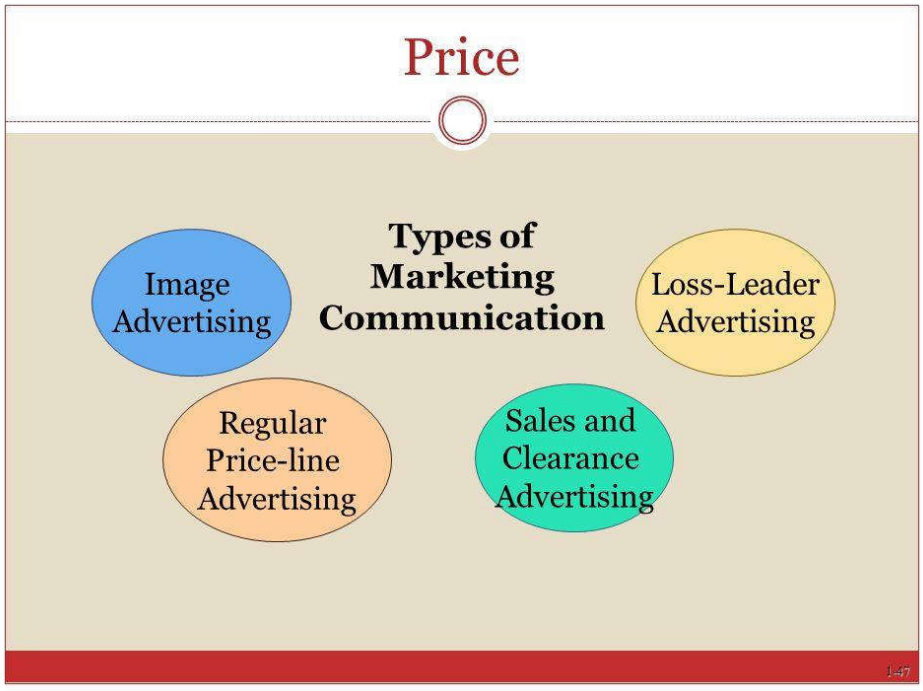
V. ADVERTISING AND INTEGRATED MARKETING COMMUNICATIONS

The following strengths of advertising: -

- Reach a mass audience stimulate large-scale demand;
- Give brand recognition
- Position the brand or product
- To expand knowledge of a particular brand
- Ensure repeatability of the treat men
- Serve as a reminder. Reveal the following weaknesses of advertising: -
- Advertising can be considered as something intrusive and, accordingly, it will try to avoid it can pollute the information environment
- It wastes a large part of its impact because of its mass orientation. Advertising should be used in the integrated marketing communications to maximize its strengths
- Taking into account other tools MK and with them.
- Advertising is very good when it builds brand image, appealing to a wide audience through the media.



VI. TYPES OF MARTKETING – COMMUNICATIONS



.

TYPE OF PROMOTION	SENDER	ENCODING	TRANSFER MECHANISM	DECODING BY RECEIVER	RESPONSE	FEEDBACK
Personal selling	Canon Office Equipment	Sales presentation on new model office copier	Canon sales representative	Office manager and employees in local firm discuss Canon sales presentation and those of competing suppliers	Order placed Canon copier	Information that for customers are reacting positively to the message
Two-for-one coupon(sales promotion)	Wendy's Hamburgers	Wendy's marketing department and advertising agency	Coupon inserted in weekend newspaper	Newspaper reader sees coupon for hamburger and saves it	Hamburgers purchased by consumers using the coupon	Information that customers are reacting positively to the message
Television advertising	Movie producer	Advertisement for a new movie is developed by the producer's advertising agency	Network television during programs with high percentage of viewers in target market	Audience sees ad but few decide to go to the movie	Small number of movie tickets purchased	Communication failed to interest and motivate the target market

VII. CONCLUSION: Marketing is the most exciting of all business sports. It is the heartbeat of every successful business. It is continually changing in response to the explosion of information, the expansion of technology, and the aggressiveness of competition, at all levels and everywhere.All business strategy is marketing strategy. Your ability to think clearly and well about the very best marketing strategies, and to continually change and upgrade your activities, is the key to the future of your business. Product is considered as main element of the marketing mix, Price is understood through the value of the product which is paid by the buyer to purchase the product. Place or distribution strategy of marketing mix is concerned to make products accessible to customers at favorable time and place. The last element of marketing mix, promotion strategy is focused to various activities like publicity, public relations, sales promotion

References:

1) *Marketing Management by – Pillai RSN*
2) *Basic Marketing Management – Rudani R.B.*
3) *Principles of Management – Philip T Kotler*
4) *Service Marketing : People Technology Strategy*
5) *Marketing 4.0 : Moving from Traditional to Digital – Philip Kotler*

A STUDY ON MUTUAL FUNDS

1)B.MOUNIKA
Assistant Professor, Department MBA,
Malla Reddy College of Engineering
Hyderabad.
Mail Id: mounikad2009@gmail.com

2) K.LAVANYA
Assistant Professor, Department MBA,
Malla Reddy College of Engineering
Hyderabad.
Mail Id: lavanyakummari.k@gmail.com

ABSTRACT:

Mutual funds mobilize the savings of the people and channelize it to the money and capital market. One of the main advantages of mutual funds over any other investment to small investor is that they give small investors access to professionally managed, diversified portfolio of equities, bonds and other securities, which is rather impossible for a small investor to create with a small amount of capital he/ she owns. Mutual funds constitute a very important component of the capital market in developed countries and are now becoming vibrant in emerging markets like India. The origin of mutual funds industry in India can be traced in the enactment of the Unit Trust of India (UTI) Act in 1963. Due to historic reasons, the UTI enjoyed the total monopoly in the initial years and until now continues to maintain the largest market share. The industry has now moved from complete monopoly to that of a monopolistic competition. Presently, the share of Net Assets of mutual funds is more than 7 percent of India's gross domestic product (GDP). Also, the monies accredited to mutual funds form an adequate part of gross domestic savings (GDS) in the country.

I.INTODUCTION:

Individual investors have developed keen interest within the capital market, attaining higher returns and capital gains in conjunction with transaction concessions. Since little or no investors typically haven't got adequate time, knowledge, expertise and resources for directly approaching the capital market, they have to believe associate intermediate that undertakes wise investment choices and provides the advantages of skilled experience. this is often what a investment company will. Mutual funds square measure dominated by SEBI. SEBI has the authority to issue pointers and to supervise and regulate the in operation of mutual funds through Mutual Funds rules, 1993 square measure amended from time to time. Mutual funds offer stability arid property to the exchange conjointly. An investment company could be a pool of investment managed professionally for the aim of buying varied securities and culminating them into a powerful portfolio which is able to supply enticing returns over and on top of the riskless returns that are presently being offered by the market. Investment company could be a money product that invests in stocks or bonds.

II. REVIEW OF LITERATURE

Literature on no depository financial institution performance analysis is massive. some analysis studies that have influenced the preparation of this paper well are mentioned throughout this section Sharpe,

C.Srinivas Yadav and Hemanth N C (Feb 2014) Have studied Performance of chosen Equity Growth Mutual Funds in India: associate Empirical Study throughout initial Gregorian calendar month 2010 To thirty initial would possibly 2013. The study evaluates performance of chosen growth equity funds in Asian nation, disbursed exploitation portfolio performance analysis techniques like Sharpe and Treynor live. S&P CNX corking has been taken as a results of the benchmark. The study conducted with fifteen equity growth Schemes (NAV) were chosen from high ten AMCs (supported AUM) for the quantity first June 2010 to thirty first could 2013(3 years).

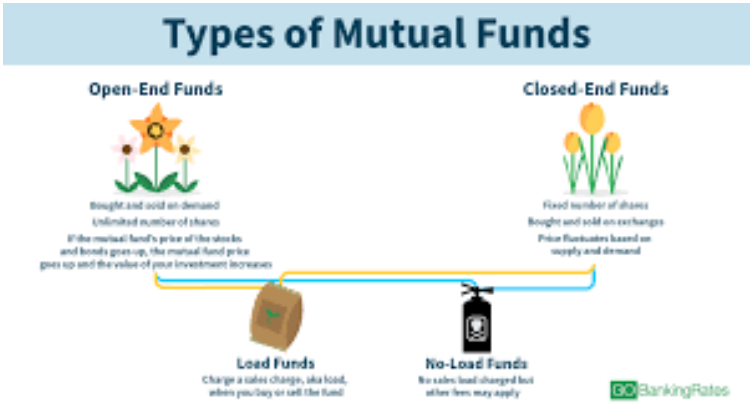
Mrs.V. Sasikala and Dr. A. Lakshmi (Jan 2014) Have studied The no depository financial organization Performance Between 2008 And 2010: Comparative Analysis. The paper entitled "comparative analysis of non depository financial institution performance between 2008 & 2010. The paper was undertaken to know the once meltdown amount risks and returns of 2008 high hundred mutual funds and compare with 2010 high hundred mutual funds open in Business recently. The analysis of alpha, beta, variance, Sharpe relation and R-squared ar declare high, low, average, on prime of average and below average of risks and square measure offered of funds.

Dr. E. Priyadarshini (2013) has done Analysis of the Performance of Artificial Neural web work Technique for prediction non depository financial institution web and Values. throughout this paper, Internet and Values of 4 Indian Mutual Funds were expected exploitation Artificial Neural Network once eliminating the redundant variables exploitation PCA and put together the performance was evaluated exploitation customary arithmetic measures like MAPE, RMSE, etc.

III. KEY TERMS:

Portfolio, Corpus, Diversified equity mutual funds, Balance fund, SIP, NAV

IV. TYPES:



1. Mutual Fund

A mutual fund is an investment security type that enables investors to pool their money together into one professionally managed investment. Mutual funds can invest in stocks, bonds, cash and/or other assets. These underlying security types, called *holdings* combine to form one mutual fund, also called a *portfolio*.

.

2. Mutual Fund Loads

Loads are fees charged to the investor when buying or selling certain types of mutual funds .There are four types of loads: **Front-end loads** are charged up front (at the time of purchase) and average around 5% but can be as high as 8.5%.

3. Mutual Fund Share Class:

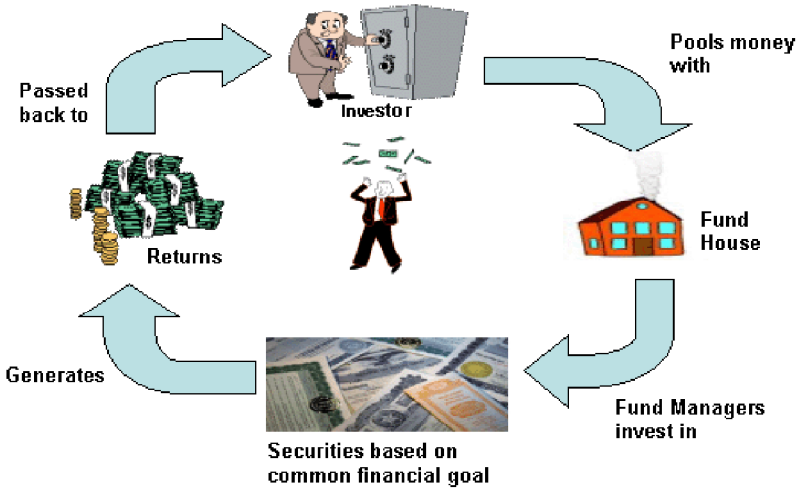
Each mutual fund has a share class, which is basically a classification of how the fund charges fees. There are several different types of mutual fund share classes, each with its own advantages and disadvantages, most of which center upon expenses.

- Class A shares are also called "front load" funds because their fees are charged on "the front" when the investor first buys shares of the fund. The loads typically range from 3.00% to 5.00%. A shares are best for investors who are using a broker and who plan to invest larger dollar amounts and will buy shares infrequently. If the purchase amount is high enough, investors may qualify for "breakpoint discounts."
- Class B Share Funds are a share class of mutual funds that do not carry front-end sales charges, but instead, charge a contingent deferred sales charge (CDSC) or "back-end load." Class B shares also tend to have higher 12b-1 fees than other mutual fund share classes.For example, if an investor purchases mutual fund Class B shares, they will not be charged a front-end load but will instead pay a back-end load if the investor sells shares prior to a stated period, such as 7 years, and they may be charged up to 6% to redeem their shares. Class B shares can eventually exchange into Class A shares after seven or eight years. Therefore they may be best for investors who do not have enough to invest to qualify for a break level on the A share but intend to hold the B shares for several years or more.
- Class C Share Funds charge a "level load" annually, which is usually 1.00%, and this expense never goes away, making C share mutual funds the most expensive for investors who are investing for long periods of time.The load is usually 1.00%. In general, investors should use C shares for short-term (less than 3 years).
- Class D Share Funds are often similar to no-load funds in that they are a mutual fund share class that was created as an alternative to the traditional and more common A share, B share and C share funds that are either front-load, back-load or level-load, respectively.

- Class Adv Share Funds are only available through an investment advisor, hence the abbreviation "Adv." These funds are typically no-load (or what is called "load-waived") but can have 12b-1 fees up to 0.50%. If you are working with an investment advisor or another financial professional, the Adv shares can be your best option because the expenses are often lower.
- Class Inst Share Funds (aka Class I, Class X, or Class Y) are generally only available to institutional investors with minimum investment amounts of $25,000 or more.
- Load-Waived Funds are mutual fund share class alternatives to loaded funds, such as A share class funds. As the name suggests, the mutual fund load is waived (not charged). Typically these funds are offered in 401(k) plans where loaded funds are not an option. Load-waived mutual funds are identified by an "LW" at the end of the fund name and at the end of the ticker symbol. For example, American Funds Growth Fund of America A (AGTHX), which is an A share fund, has a load-waived option, American Funds Growth Fund of America A LW (AGTHX.LW).
- Class R Share Funds do not have a load (i.e. front-end load, back-end load or level load) but they do have 12b-1 fees that typically range from 0.25% to 0.50%. If your 401(k) only provides R share class funds, your expenses may be higher than if the investment choices included the no-load (or load-waived) version of the same fund.

4. Expense Ratio

Even if the investor uses a no-load fund, there are underlying expenses that are indirect charges for use in the fund's operation. The expense ratio is the percentage of fees paid to the mutual fund company to manage and operate the fund, including all administrative expenses and 12b-1 fees. The mutual fund company would take those expenses out of the fund prior to the investor seeing the return. For example, if the expense ratio of a mutual fund was 1.00%, and you invested $10,000, the expense for a given year would be $100. However, the expense is not taken directly out of your pocket. The expense effectively reduces the gross return of the fund. Put differently, if the fund earns 10%, before expenses, in a given year, the investor would see a net return of 9.00% (10.00% - 1.00%).



5. Index Funds

An index, with regard to investing, is a statistical sampling of securities that represent a defined segment of the market. For example, the S&P 500 Index, is a sampling of approximately 500 large capitalization stocks. Index funds are simply mutual funds that invest in the same securities as its benchmark index. The logic in using index funds is that, over time, the majority of active fund managers are not able to outperform the broad market indexes. Therefore, rather than trying to "beat the market," it is wise to simply invest in it. This reasoning is a kind of "if you can't beat 'em, join 'em" strategy. The best index funds have a few primary things in common. They keep costs low, they do a good job of matching the index securities (called tracking error), and they use proper weighting methods. For example, one reason Vanguard has some of the lowest expense ratios for their index funds is that they do very little advertising and they are owned by their shareholders. If an index fund has an expense ratio of 0.12 but a comparable fund has an expense ratio of 0.22, the lower cost index fund has an immediate advantage of 0.10. This only amounts to only 10 cents savings for every $100 invested but every penny counts, especially in the long run, for indexing.

6. Market Capitalization

With investment securities market capitalization (or market cap), refers to the price of a share of stock multiplied by the number of shares outstanding. Many equity mutual funds are categorized based on the average market capitalization of the stocks that the mutual funds own. This is important because investors need to be sure of what they are buying. *Large-cap Stock Funds* invest in stocks of corporations with large market capitalization, typically higher than $10 billion. These companies are so large that you have probably heard of them or you may even purchase goods or services from them on a regular basis. Some large-cap stock names include Wal-Mart, Exxon, GE, Pfizer, Bank of America, Apple and Microsoft. *Mid-cap Stock Funds* invest in stocks of corporations of mid-size capitalization, typically between $2 billion and $10 billion. Many of the names of the corporations you may

recognize, such as Harley Davidson and Netflix, but others you may not know, such as SanDisk Corporation or Life Technologies Corp. *Small-cap Stock Funds* invest in stocks of corporations of small-size capitalization, typically between $500 million and $2 billion. While a billion-dollar corporation may seem large to you, it's relatively small compared to the Wal-Marts and Exxons of the world. A subset of small-cap stocks is "Micro-cap," which represents mutual funds investing in corporations with average market capitalization usually less than $750 million.

7. Mutual Fund Style

In addition to capitalization, stocks, and stock funds are categorized by style which is divided into Growth, Value or Blend objectives. *Growth Stock Funds* invest in growth stocks, which are stocks of companies that are expected to grow at a rate faster than the market average. *Value Stock Funds* invest in value stocks, which are stocks of companies that an investor or mutual fund manager believes to be selling at a price lower than the market value. Value Stock Funds are often called Dividend Mutual Funds because value stocks commonly pay dividends to investors, whereas the typical growth stock does not pay dividends to the investor because the corporation reinvests dividends to further grow the corporation. *Blend Stock Funds* invest in a blend of growth and value stocks. Bond funds also have style classifications, which have 2 primary divisions:

1) Maturity/Duration, which is expressed as long-term, intermediate-term, and short-term,

2) Credit quality, which is divided into high, investment grade, and low (or junk).

8. Balanced Funds

Balanced Funds are mutual funds that provide a combination (or balance) of underlying investment assets, such as stocks, bonds, and cash. Also called hybrid funds or asset allocation funds, the asset allocation remains relatively fixed and serves a stated purpose or investment style. For example, a conservative balanced fund might invest in a conservative mix of underlying investment assets, such as 40% stocks, 50% bonds, and 10% money market.

9. Target Date Retirement Funds

This fund type works like its name suggests. Each fund has a year in the name of the fund, such as Vanguard Target Retirement 2055 (VFFVX), which would be a fund best suited for someone expecting to retire in or around the year 2055. Several other fund families, such as Fidelity and T. Rowe Price, offer target date retirement funds. Here's basically how they work, other than just providing a target date: The fund manager assigns a suitable asset allocation (mix of stocks, bonds, and cash) and then slowly shifts the holdings to a more conservative allocation (fewer stocks, more bonds, and cash) as the target date draws closer.

10. Sector Funds

These funds focus on a specific industry, social objective or sector such as health care, real estate or technology. Their investment objective is to provide concentrated exposure to specific industry groups, called sectors. Mutual fund investors use sector funds to increase exposure to certain industry sectors they believe will perform better than other sectors. By comparison, diversified

mutual funds--those that do not focus on one sector--will already have exposure to most industry sectors. For example, an S&P 500 Index Fund provides exposure to sectors, such as healthcare, energy, technology, utilities, and financial companies. Investors should be careful with sector funds because there is increased market risk due to volatility if the sector suffers a downturn. Over-exposure to one sector.

Recommendations – Companies tie up:-

- Aditya Birla Sun Life Frontline Equity **Fund** (G)
- SBI Bluechip **Fund** G.
- Mirae Asset Emerging Bluechip **Fund** G.
- L&T India Value **Fund** (G)
- Kotak Select Focus **Fund** (G)
- ICICI Prudential Equity and Debt **Fund** G.
- ICICI Prudential Bluechip **Fund** G.
- Reliance Small Cap **Fund** (G)

CONCLUSION: Mutual funds are a popular investment avenue among investors, as they are easy to invest in and give higher returns as compared to other traditional asset classes such as FDs or saving bank deposits. Mutual fund business is one all told the fastest growing business in Republic of India and it's already established in foreign countries. Finance in mutual funds might be a heap of safe as compared to equity likewise as a result of it offer handsome returns. Supported differing types of equity and debt mutual funds the following conclusion area unit usually created.

References:

- *A Study of Mutual Funds, House Report #2247 [August, 19621, Report of the Committee on interstate and Foreign Commerce.*
- *"The Performance of Hedge Funds: Risk, return and Incentives," The Journal of Finance, 54, 833-874 Author: Ackermann, C., R. McEnally, and D. RavenscraftPublished: 1999*
- *Money Control web site*
- *Common sense on Mutuval Funds – New Imperatives for the intelgent investors - John C. Bogle*

DECISION MAKING IN PLANT LAYOUT – PRODUCTION

GLN SRAVAN KUMAR
Assistant Professor,
Malla Reddy College Of Engineering,MRCE, Maisamaaguda, Hyderabad 500100
e-Mail: sravanhrm21@gmail.com

I. ABSTRACT: The study seeks to a Plant location refers to the choice of optimal place which is best or favorable at which production plant can be established such that efficiency and effectiveness can be maximized. It was decided by entrepreneur before establishing a new enterprise. The plant Layout refers to the physical arrangements of production facilities. It is the configuration of various departments work centers and equipments in the conversion process, this is floor plan of physical facilities, which are used in the production, these process depends in functional conditions, techniques, experience and scope of performance in future concern. It is a choice of the region where the men, materials, money, machines and equipments are bought to forward together for setting up a factory or business. Productivity is defined as a total output per one unit of a total input. In control management, productivity is a measure of how efficiently a process runs and how effectively it uses resources.

II. KEY WORDS: Operation Management, Layout, Facilities, Technology Obsolescence, Value Added Product / Services, Sub System, Buy or Make decision, Productivity

III. INTRODUCTION: Organization aims to promote its product among customers which helps to obtain substantial sales order, the term is communicated to the production subsystem which is concerned with the management of physical resources for the production of an item or provision of service. To manufacture a product as per the specifications, the production functions needs to organize its resources like raw materials, equipments, labour and working capital according to the predetermined production plan. Technology obsolescence may force some industries to use the phase in strategy for introducing the next model of same product or service to retain and improve its market segment. It is amply clear that all functional subsystem of any business organization are interwoven by many linkages. Production and Operations Management is the process which combines and transforms various resources used in the production / operation sub system of an organization into value added product / services in a controlled manned as per the polices of the organization. The set of

interrelated management activities which are involved in manufacturing certain products is called Production management.

IV. MATERIALS CLASSIFICATION:

For the selection of a location to start a production was divided into two types:-

(A) Ubiquities: The Raw materials include water, air, sand etc which is available everywhere which helps in the operational research by using " Model Driven Architecture" which contains core theme on " How to capture each kind of knowledge in a form that transferable , reusable, teachable and deployable."

(B) Localized Materials: The raw materials that are found in some specified locations

Ex: - Iron, coal, Bamboo, Wood etc

It was divided into two types

- Pure Materials
- Weight lose Materials

V. PRINCIPLES OF LAYOUT:

1. The Principle of minimum travel
2. The Principle of sequence
3. The Principle of usage
4. The Principle of compactness
5. The Principle of safety and satisfaction
6. The Principle of flexibility
7. The Principle of minimum investment

VI. FACTORS AFFETING PLANT LAYOUT:

→ Policy Management
→ Plant Location
→ Nature of the Product
→ Volume of Production
→ Geographical constraint – Transportation
→ Availability of floor space
→ Nature of Manufacturing process
→ Repair equipments and availability
→ Economical situation & Country constraint
→ Availability of Infra Structure

VII TYPES OF INDUSTRIES:

(A) Synthetic: Initiate from a Natural Product

In these type of industries several elements under go in production process to form a finished products (Two or more materials to form an output)

Ex: Chemical and Paper industry

(B) Analytical : Systematic, logical, methodology

It involves in conversion of raw material into different elements, raw material divided and sub divided into several useful units

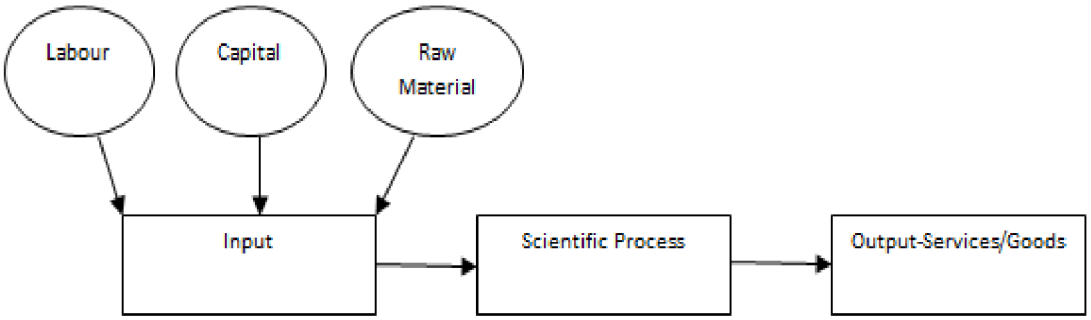
Ex: Oil refinery – Separates crude oil into Kerosene, diesel, petrol

(C) *Conditioning:* These industries to produce products with variable physical properties

Ex: Metal working industry

(D) Extractive : Such industries extract one element from the group of elements

Ex: Metal from ore



VIII LOCATION BASED ANALYSIS METHODS:

(A) Line Balance Procedure Method:

Line of **Balance** (LOB) **Line** of **Bal ance** (LOB) is a management control **process** for collecting, measuring and presenting facts relating to time (see Schedule Control), cost and accomplishment – all measured against a specific plan.

$$\mathbf{Xa} = \frac{\text{Sum of the task time}}{\text{Cycle time}}$$

(B) Centroid Method:

This method helps in business to identify the location of facilities used for manufacturing and other product related aspects. It is also called as Agglomerative Clustering Method.

$$X = \frac{\text{Sum of volume X Sum of Coordinators}}{\text{Sum of Volume per day}}$$

(C) Location Based Scoring Method:

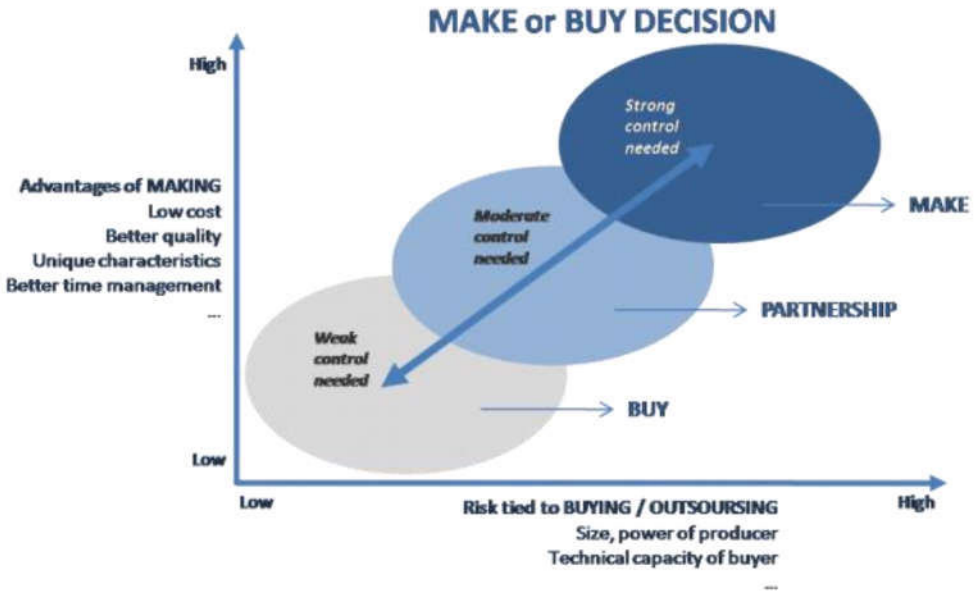
The scoring model is a quantitative analysis method that reveals relationships between several, difficult to compare key figures for locations, regions or areas. The goal is to derive indicators to aid in decisions regarding each location or area. These criteria in search with availability of Infra structure, Size of Market, Tax benefit concession – geographical constraints, availability of labour and resources etc.

IX. ECONOMIC ORDER QUANTITY:

Buy Opinion

It is very first step in process planning and it is the deciding factor whether the product or service made in an organization or should brought from others The make-or-buy decision is the act of making a strategic choice between producing an item internally (in-house) or buying it externally (from an outside supplier Issues like government regulation, competing firms, and market trends all have a strategic impact on the make-or-buy decision.

Make Opinion



Buy Decision = Q1= Economic order Quantity = $\sqrt{\frac{2C_oD}{C_c}}$

Total Cost = $D \times P + \frac{DCo}{Q1} + \frac{Q1 \times Cc}{2}$

Make Opinion = Q2= Economic Order Quantity = $\sqrt{\frac{2SD}{Cc(1-D/R)}}$

Total Cost = $D \times P + \frac{D \times S}{Q2} + Cc\,(R - D)\,\frac{Q2}{2 \times R}$

Where

D = Demand in the Year / Month/ Quarter

P = Purchase Price per Unit

Cc = Carrying Cost per unit in Year / Month/ Quarter

Co = Ordering Cost / Order

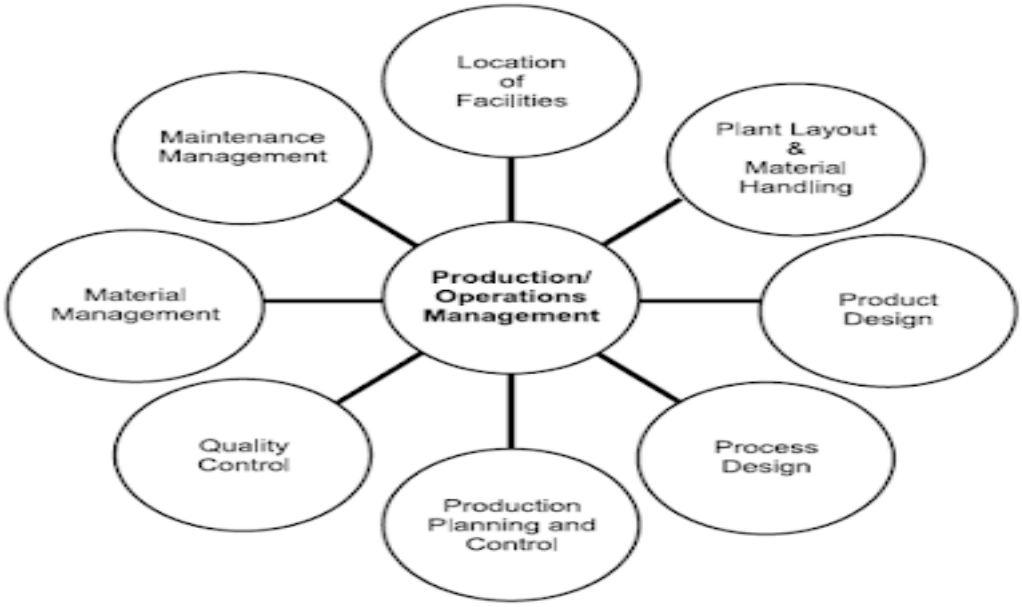
S = Set up time

EOQ = Economic Order Quantity

R = Production rate

Q1 = Economic Order Size – Buy

Q2 = Economic Production Size - Make



XI. RECOMMENDATIONS:

- Choosing the geographical constraint plays an important role in Production company
- Transportation is the criteria to make time and money saving process which was interlinked to geographical section
- Choice of owner depends on Buy or Make decision which provides result after the production
- Working out the decision making regarding Buy or Make should be performed with perfect margins then the industry will have best financial strength
- Grabbing familiar idea regarding Locating the problems (political, economical) issue (legal) , availability of labour (skilled and semi skilled), reducing the cost of production.
- The Company has indicated a low inherent risk regarding the appropriateness of the location of the flow meters whereas the verifier deems the risk to be higher requiring more robust control activities such as maintenance or better location on the inlet to the emissions sources.
- The `process activity` for enhancing human resource for the on board ship`s staff with regards to competency and training was found to be in sufficient.

X. CONCLUSION :

Choosing the best location for a which to apt a production / manufacturing industry it is crystal clear that to workout 360 degrees monitoring like transportation, labour availability, decision regarding make or buy a material to make product into finished goods. Effective and Efficient¬Strategic Implications regarding Competitive Priorities (regarding capacity, process, flexibility, cost, quality, customer, image. Layout will meet Firm's competitive requirements, which may results for customer relation management.

XII. REFERENCES:

1. *Printing plant layout and facilities design – A. John Gels*
2. *Production and operations Management 2nd edition – R. Panneerselvam*
3. *Operations Management – Aswarthaappa*
4. *Plant Layout and Material handling – James M .Apple.*
5. *Process Plant Instrumentation Design & Upgrade – Miguel J. Bagajewicz*

Gender Discrimination and its Impact on Employee Performance at Work Place

S. PRASHEELA
Asst.Professor, Department of Business Administrations,
MRCE – MBA, Hyderabad.
Mail id: s.prasheela1990@gmail.com

Abstract : This research paper explores the issue of gender discrimination in workforce and its impact on the satisfaction and motivation, commitment and enthusiasm and stress level of employees. Close ended questionnaire was administered from 526 males and females oflower, middle and higher category employees of public and private health and education departments of Hyderabad . Gender discrimination in workforce was measured through independent samples-t test. The analysis shows that females were discriminated more than males in private organizations. Thus the findings show that females are discriminated more than males in private sector than in public sector. The impact of gender discrimination on satisfaction &motivation, commitment & enthusiasm and stress level was assessed through Pearson product moment correlation co-efficient. The results show that gender discrimination decreases satisfaction & motivation and commitment & enthusiasm level of employees, and increases the stress level in the employees.

Key words: Gender discrimination, Workforce, Satisfaction & Motivation, Commitment & Enthusiasm, Stress Level.

I .Introduction :

In (1998 Decenzo,) has explained that HRM in any organization is concerned with the staffing, training, developing, motivating and maintaining the employees. Staffing is concerned with recruiting and selection ,strategic human resource planning,. , HRM, Fifty years ago was considerably simpler because work force was strikingly homogenous. Now-a-days work force scenario has changed has changed from homogeneous to heterogeneous type.Managing diversity is very essential for any organization, especially in this era of globalization. Managing diversity is required to close the unfair discrimination and thus enable employees to compete on equal basis.

II . Literature Review :

Researchers found that women in Sweden suffer more from sticky floor effects than glass ceilings. Their study also showed that women with small children face a largest gender penalty in careers. Gender penalty is larger for younger and older women and less for middle aged women. There was no any empirical support in their study that women have lesser career opportunities in the private sector than in public sector, relative to men. Susan et al (1998) have focused on the work place gender discrimination rational bias theory. According to this theory, decision makers may choose to discriminate if they believe that their superiors or others having power over their careers expect or prefer it. The findings of their research showed that businessmen discriminated women and people at the top of the organization are most biased against women than people at the bottom. Their study has also confirmed that management support discrimination, though those discriminations were less than the findings of earlier research, reflecting increasing equal opportunity. It was also confirmed through their research that the discrimination is more because of external pressures than from internal. Habib (2000) has studied the effects of Brick Wall and Glass ceiling in public administration of Bangladesh. His analysis has shown that women are discriminated in civil services of Bangladesh from entry to the higher posts (Glass Ceiling). Social cultural factors are the principal stumbling blocks and build a wall for entry of woman into civil services. Their career path was hindered by the impediments of the systematic and attitudinal reasons. Government laws and regulations in this regard are proving ineffective. However, that discrimination was not for the women who came from upper class. This disparate treatment against women had implications for their morale, motivation and performance. Uzma (2004) found out that identity is created through the society, environment and parents. It is a two-way process - how people view you and how you view yourself. Attitude of parents towards their children formulate their identity. Parents usually consider their daughters as weak, timid, and too vulnerable; they need to be protected by the male members of the society. Because of this reason females cannot suggest or protest. This is the first step of subjugation and suppression. According to her; even the educated females have the double identity – professional and private. Another finding of her research was that the income of the women is not considered as the main financial source for the family, but as supplementary to the income of their males. She also found that those results were not valid for the upper and advanced families, where complete freedom is given to their females.

Research Methodology: The study objective is to add few factors and dimensions of gender discrimination which are responsible for the effect on employee performance/Productivity. Here the research will present the relationship between gender discrimination and employee productivity in

Telecom sector of Pakistan, narrow down to the telecom of telenor and Ufone of Islamabad..

Independent variables :

Independent variables affect the dependent variables either negatively or positively, it depends the circumstances and vary to vary in different studies. Sex(Gender) discrimination and its dimensions are the independent variable. Gender discrimination is one of the leading social problems all over the world.

Dependent variable:

The concerned study variable is known as dependent variable which may be affected/influenced by the other independent variables and can be described the variation of discrimination over the productivity of the organization. Productivity is the major variable (dependent).

III.. Gender Discrimination :

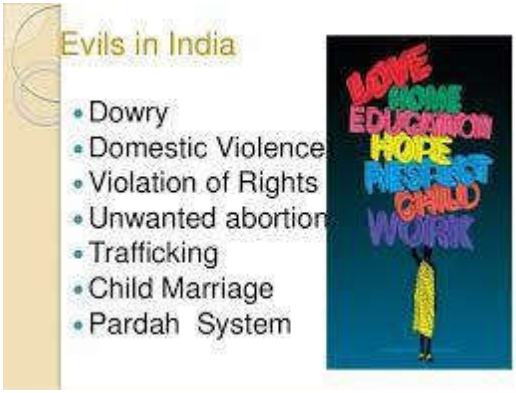
No law has ever attempted to define precisely the term' discrimination said Wayne (1995', in the context of workforce, it can be defined as the giving of an unfair advantage (or disadvantage) to the members of the particular group in comparison to the members of other group. Narrating the decisions of the courts, Ivancevich (2003) writes that in interpreting title VII of the 1964 Civil Rights Act and other laws, the United States' courts have held that both intentional (disparate treatment) and unintentional (disparate impact) acts of covered entities may constitute illegal employment discrimination.

IV. Statement of the Problem :

Gender discrimination turns the employees emotionally brittle, simple peace loving employees transform into paranoid and suspicious, fearful and angry individuals. Elimination of Gender

Discrimination is crucial for the satisfaction and motivation, commitment and enthusiasm and less stress of the employees.

V. Problems in India: The Marriage Market



VI. Dowry:

The custom of dowry in Indian marriages is a deep-seated cultural phenomenon that has been described as one of the largest obstacles to "confront India on her road to economic and social justice. The custom is held responsible for a number of problems perpetrated against the nation's women, including dowry violence, bride burning, and wife murder. I argue that dowries exist because of a combination of two reasons. First, there is an excess supply of women in the Indian marriage market that results in the use of dowry as an equilibrating mechanism. Secondly, a differential in the patterns of human capital accumulation of men and women have led to a larger positive benefit from marriage for women than for men, the net difference of which is theoretically equivalent to the amount of the dowry. Both these explanations for the existence of dowry are fundamentally.

VII. Domestic violence:

According to a National Family and Health Survey in 2005, total lifetime prevalence of domestic violence was 33.5% and 8.5% for sexual violence among women aged 15–49.[A 2014 study in The Lancet reports that although the reported sexual violence rate in India is among the lowest in the world, the large population of India means that the violence affects 27.5 million women over their lifetimes. However, A survey carried out by the Thomson Reuters Foundation ranked India as the most dangerous country in the world for women.

VIII. National Crime Records Bureau says reported that reported crime rate of 46 per 100,000, rape rate of 2 per 100,000, There are several domestic violence laws in India. The earliest law was the Dowry Prohibition Act 1961 which made the act of giving and receiving dowry a crime. In an effort to bolster the 1961 law, two new sections, Section 498A and Section 304B were introduced into the Indian Penal Code in 1983 and 1986. The most recent legislation is the Protection of Women from Domestic.

IX. Discussion and Conclusion:

Employee (male/female) refers to the personnel who executes different functions, duties and responsibilities in order to achieve the organizations/institutes objectives. Organizational productivity and performance affected by employee performance and employee performance affected by gender discrimination. The survey conducted in telecom sector especially ufone and telenor companies‟ employees male and female in Islamabad for the research study to find out the relationship and affect on employee productivity. Therefore statistical tools employed correlation and regression analysis and model summary shows the significant relationship between gender discrimination and employee productivity. All hypotheses proved and claimed about significant association and relationship between gender discrimination & employee productivity. Gender differences in hiring, gender biasness in promotion and gender disparity in provision of goods and facilities have negative relationship with productivity.

The human resource managers should be careful while hiring & promotion of the employees and providing facilities to employees in order to avoid any gender discrimination because it has a direct relationship on employee productivity and which will reduce organizational productivity. The human resource managers should provide facilities and provision of goods according to the ratio of men and women. In the mean while promotion of employees on merit and seniority wise. They must accept the importance of female workers and realize the international scenario of women fair representation and quotas in hiring, and facilities in organizations, otherwise they will get defame and bad reputation in corporate sector of the world which will reduce international business cooperation and lose large number of business orders. The decline in number of business orders from all over the world will affect the profitability and survival of the organizations.

Therefore, top management must avoid discriminatory practices in hiring, promotion, and facilities to the employees irrespective of what is his/her gender. Positive and loyal culture will build in the organization which will increase the productivity and profitability of the organization.

References:

- *Fanslow, J., Robinson, E., Crengle, S., & Perese, L. (2010). Juxtaposing beliefs and reality: Prevalence rates of intimate partner violence and attitudes to violence and gender roles reported by New Zealand women. Violence against Women, 16(7), 8128-8131.*
- *Ridgeway, C., & Correll, S. (2004). Unpacking the gender system: A theoretical perspective on gender beliefs and social relations.* Gender and Society, 18
- *Good and Mad: The Revolutionary Power of Women's Anger by Rebecca Traister*
- *We should all be feminist by Chimamanda Ngozi Adichie*
- *I Am Malala: The Story of the Girl Who Stood Up for Education and Was Shot by the Taliban by Malala Yousafzai*

TALENT MANAGEMENT

LAXMI PRASANNA,
Student – Department of MBA,
Malla Reddy College of Engineering
Mail Id: prasuprasanna2989@gmail.com

KAVYA
Student – Department of MBA,
Malla Reddy College of Engineering
Email id: kavya.reddy.kr941@gmail.com

ABSTRACT:

This paper observes that study of entering a new era: the emergence of "Talent Management." While strategic HR continues to be a major focus, HR and L&D organizations are now focused on a new set of strategic issues: How can we make our recruiting process more efficient and effective by using "competency-based" recruiting instead of sorting through resumes. Talent management is also known as HCM (Human Capital Management), HRIS (HR Information Systems) or HRMS (HR Management Systems), and HR Modules. Organizations are made up of people: people creating value through proven business processes, innovation, customer service, sales, and many other important activities. As an organization strives to meet its business goals, it must make sure that it has a continuous and integrated process for recruiting, training, managing, supporting, and compensation acting these people. This paper strives to understand the Talent Management as an emerging concept and seeks to understand its objectives, process, and critical success factors.

I INTRODUCTION:

The 21st century brings a unique and unprecedented set of challenges and potential opportunities for organizations. The competitive landscape is increasingly global in scope. Today, even small or midsize companies are challenged to compete on a global scale. The number of multinational corporations has doubled since 1990 . Growth opportunities are more prevalent in developing regions of the world where the middle class and new consumers are emerging rapidly. In fact, the global middle class is expected to grow by 172 percent from 2010 to 2030. At the same time, there are shifts in the available talent pool and workforce demographic. An aging demographic in the West is yielding to higher proportions o f college graduates coming from Latin America, Russia, and Asia. Individuals are being asked to do more with less, but they also have higher expectations for their employers and are more mobile than previous generations. Consumers also have higher expectations, and the companies that make the best use of technology and data analytics to better understand their customers and break down barriers will win . The pace of market growth, the speed of innovation, and ever - changing demographics create more

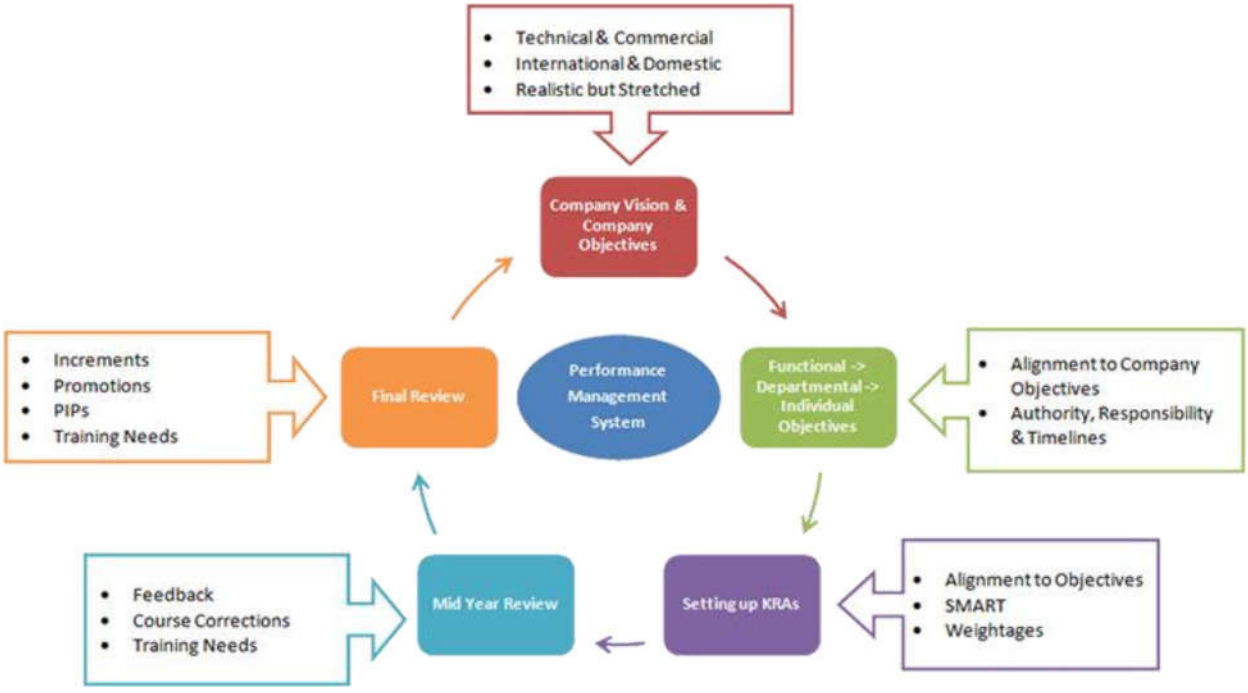
opportunities to both create and extract value, but its often more difficult to pinpoint those opportunities and act upon them. So, how do companies compete in this increasingly complex and changing environment? One of the central differentiators for companies is a sound business strategy. Leaders who are adept at creating winning strategies, make an integrated set of choices regarding issues such as where the business will be active, how to get there, how to win, the specific actions and moves, and the model of profit generation . A business strategy articulates the direction a business will pursue and the steps it will take to achieve its goals. A good strategy has at least three characteristics: it creates value, it‟s unique, and it cannot be imitated

II. OBJECTIVES OF STUDY:

1. To assess the role of Talent Management in Industry.

2. To analyze the Talent Management Process

3. To find out the critical success factors of Talent Management.



III. OBJECTIVES OF TALENT MANAGEMENT

i. ***Recruiting the Talent:*** The actual process of hiring starts from here. This is the stage when people are invited to join the organization.

ii. ***Selecting the Talent:*** This involves meeting with different people having same or different qualifications and skill sets as mentioned in job description. Candidat es who qualify this round are invited to join the organization.

iii. ***Training and Development:*** After recruiting the best people, they are trained and developed to get the desired output.

iv. ***Retention***: Certainly, it is the sole purpose of talent management process. Hiring them does not serve the purpose completely. Retention depends on various factors such as pay package, job specification, challenges involved in a job, designation, persona l development of an employee, recognition, culture and the fit between job and talent.

v. ***Promotion:*** No one can work in an organization at the same designation with same job responsibilities. Job enrichment plays an important role.

vi. ***Competency Mapping:*** Assessing employees‟ skills, development, ability and competency is the next step. I f required, also focus on behavior, attitude, knowledge and future possibilities of improvement. It gives you a brief idea if the person is fir for promoting further.

vii. ***Performance Appraisal:*** Measuring the actual performance of an employee is necessary to identify his or her true potential. It is to check whether the person can be loaded with extra responsibilities or not.

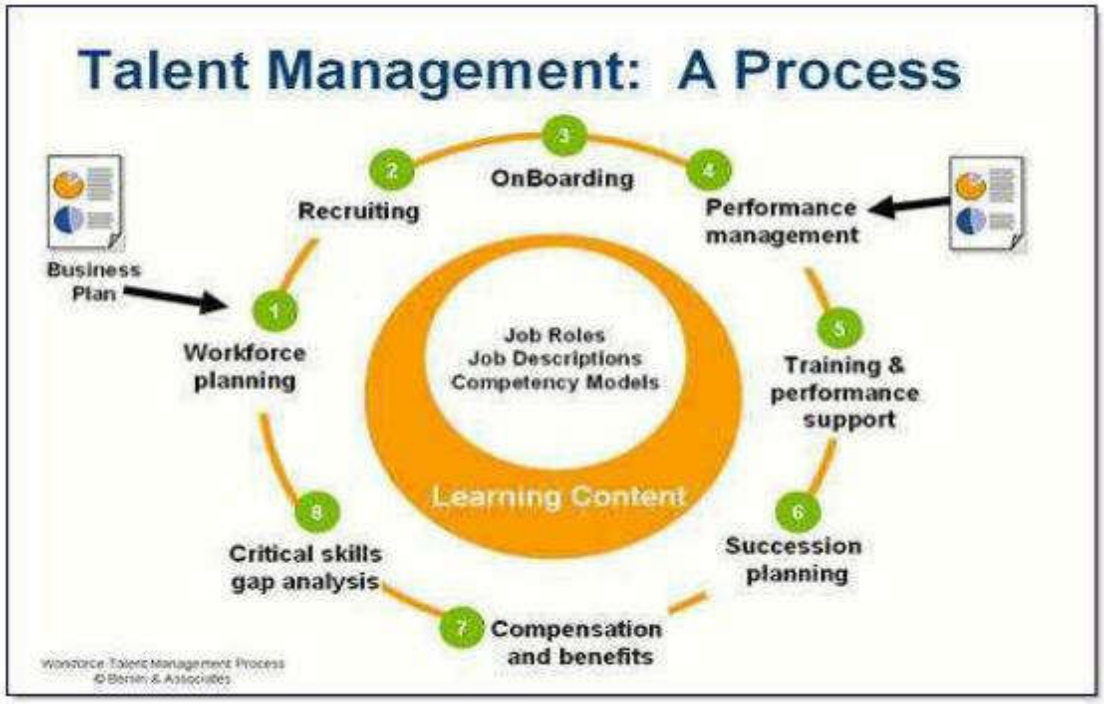
viii. ***Career Planning:*** If the individual can handle the work pressure and extra responsibilities well, the management needs to plan his or her career so that he or she feels rewarded. It is good to recognize their efforts to retain them for a longer period of time.

ix. ***Succession Planning:*** Succession planning is all about who will replace whom in near future. The employee who has given his best to the organization and has been serving it for a very long time definitely deserves to hold the top position. Management needs to plan about when and how succession will take place.

x. ***Exit:*** The process ends when an individual gets retired or is no more a part of the organization

IV. TALENT MANAGEMENT PROCESS

Organizations are made up of people creating value through proven business processes, innovation, customer service, sales, and many other important activities. As an organization strives to meet its business go als, it must make sure that it has a continuous and integrated process for recruiting, training, managing, supporting, and compensating these people.



Workforce Planning : Integrated with the business plan, this process establishes workforce plans, hiring plans, compensation budgets, and hiring targets for the year

Recruiting : Through an integrated process of recruiting, assessment, evaluation, and hiring the business brings people into the organization.

On Boarding: The organization must train and enable employees to become productive and integrated into the company more qu ickly.

Performance Management: By using the business plan, organization establishes processes to measure and manage employees. This is a complex process in itself.

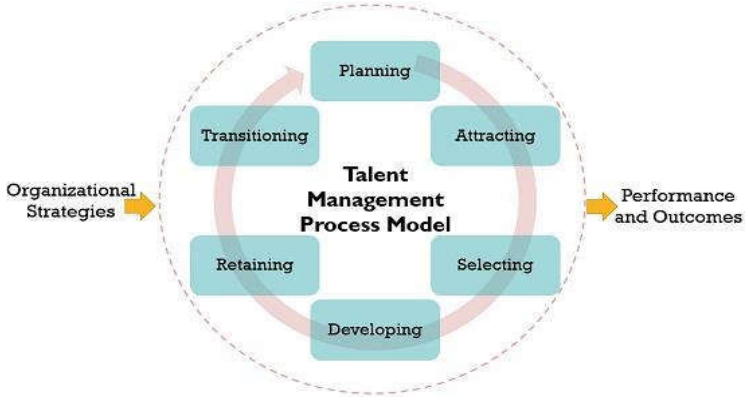
Training and Performance Support: Of course this is a critically important function. Here we provide learning and development.

Succession Planning: As the organization evolves and changes, there is a continuous need to move people into new positions. Succession planning, a very important function, enables managers and individuals to identify the right candidates for a position. This function also must be aligned with the business plan to understand and meet requirements for key positions 3-5 years out. While this is often a process reserved for managers and executives, it is more commonly applied across the organization.

Compensation and Benefits: Clearly this is an integral part of people management. Here organizations try to tie the compensation plan directly to performance management so that compensation, incentives, and benefits align with business goals and business execution.

Critical Skills Gap Analysis: This is a process we identify as an important, often overlooked function in many industries and organizations. While often done on a project basis, it can be —businesscritical." For example, today industries like the Federal Government, Utilities, Telecommunications, and Energy are facing large populations which are retiring. How do you identify the roles, individuals, and competencies which are leaving? What should you do to fill these gaps? We call this "critical talent management" and many organizations are going through this now.

Critical Success Factors of Talent Management Strategy Alignment to Corporate Strategy Ensuring that the talent strategy is closely aligned with the corporate strategy must be a priority. In developing a talent strategy, both internal and external factors should be taken into account, including fact ors influencing talent management. Strategic analysis from the business perspective should fed into an HR forecast which can help shape an organization„s tailored approach to talent management. Often organizations are at different stages of talent manageme nt strategy development. Research suggests that for a number of organization„s there was a desire to progress through the stages highlighted in following figure making the transition to a more formal, strategically-integrated approach.



Inclusive versus Exclusive Approaches: Some organizations adopt an inclusive approach to talent manage ment creating a ―whole workforce approach to engagement and talent development, while others develop a more exclusive focus segmenting talent according to need. Regardless of which approach organizations adopt, fairness and consistency must be applied in a talent management processes. Diversity considerations must also be built -into talent management processes to ensure that organizations are able to draw from the widest pools of talent possible. Involving the Right People Careful considerations need to be paid to involving the right stakeholders in the talent management strategy. Clearly, HR specialists have an important role to play in providing support and guidelines in the design and development of approaches to talent management that will fit the needs of the organization.

CONCLUSION: Talent management in an organization aims at ensuring employee recruitment, training and development, performance reviews and their compensation. Working towards enhancing a good talent management system in the organization ensures these components of human resource contribute to the success of the organization. The advantages that the components bring to the organization also outweigh the disadvantages considering organizations benefit from these approaches. These ensure the organization attracts highly qualified employees and finds it easy to retain them and hence improving their human resource element. Talent management enhances reviews that prove vital in developing employees. They reveal employee weaknesses and result in the development of training needs and programs that will improve the skills of the employees hence maintaining their talents. Employee talents also develop and change with changes in the organization needs hence increasing and improving their ability to execute their roles. Therefore, the application of talent management proves an ideal approach in employee development and improving the performance of each personnel.

REFERENCES:

1) *Talent Management – A Complete guide 2019 edition*
2) *Talent is Over rated – Geoff Colvin – 2nd edition*
3) *Human Resource Management – Aswarthapaa*
4) *Organization behavior – LM Prasad*
5) *Cite HR Website*

A STUDY ON ASSET ALLOCATION OF MUTUAL FUNDS

1) G Priyanka
Assistant Professor, MRCE,
Department of MBA,
Mail ID: priyanka.gurram2015@gmail.com

2) **D. Karunker**, Assistant Professor,
MRCE, Department of MBA
d.dumpalakarunkar@gmail.com

ABSTRACT

An investment is a commitment of funds made with the expectation of some return in the form of capital appreciation. Different investment avenues are available to the investors such as fixed deposits, insurance, post office savings/ national savings certificate, gold/e-gold, bonds, public provident fund (PPF), real estate, shares, commodities, etc. Mutual fund is one of the important investment vehicle that offer good investment prospects to the investors. Mutual fund is a trust that pools the savings of various individuals by issuing units to them and then invests it in various securities such as shares, debentures and bonds as per the stated objectives of the scheme. Further, this investment revenue offers several benefits to the investors as diversification, professional fund management, liquidity, transparency etc. Today a wide variety of mutual fund schemes are available for the investors such as Open-ended, Close-ended, Interval, Growth, Income, Balanced, Equity Linked Saving Schemes (ELSS) and Exchange Traded Funds (ETF), etc. These schemes are catering to the investors' needs, risk and return tolerance

I. KEYWORDS: Asset Allocation, Mutual Funds, Maximum Returns, Minimum Risk

II.INTRODUCTION:

The theory is that the capitalist will reduce risk as a result of every plus category incorporates a totally different correlation to the others; once stocks rise, as an example, bonds usually fall. At a time once the exchange begins to fall, land could begin generating above-average returns

The amount of AN investor's total portfolio placed in every category is set by An plus allocation model. These models ar designed to replicate the private goals and risk tolerance of the capitalist.

Any additional, individual plus categories is sub-divided into sectors (for example, if the plus allocation model needs four-hundredth of the full portfolio to be invested with in stocks, the portfolio manager could advocate totally different allocations at intervals the sector of stocks, like recommending a particular share in large-

capitalization, mid-cap, banking, producing, etc.)

III. OBJECTIVES OF THE STUDY

- To minimize volatility and. maximizing returns
- The method involves dividing your cash among classes that don't all reply to a similar economic process within the same means at a similar time.
- To determined supported your age, lifestyle, goals and risk taking appetency.
- For example a conservative capitalist are told to carry five hundredth in equity mutual funds, forty fifth in debt mutual funds and five-hitter in gold funds.

.

IV. SCOPE OF THE STUDY

This project includes the ways of study risk come analysis, applied math techniques like risk, average come, variance, Variance, Sharpe quantitative relation, Turn over quantitative relation, plus allocation are used for the analysis.

V. RESEARCH METHODOLOGY

Primary information:

The info that is directly collected from the businesses

Secondary information:

Within the gift project work {the information the info the information} has been collected from pronto on the market sources that's secondary data like websites, newspapers journals, articles etc.

VI. REVIEW OF LITERATURE

Cummings J.R (2016) Examined the link between fund size and performance for two major superannuation business sectors in Australia: retail and not-for-profit, using a particular but confidential information. Results advocate that members show pride in being endowed in larger superannuation funds for three reasons: (i) larger not-for-profit funds supply diversification edges of investment in extra and classes likewise as unlisted property and private equity, (ii) larger funds in every sectors avoid the dimensions diseconomies in investment returns documented in studies of equity Mutual Funds and (iii)larger funds build substantial savings by spreading fastened in operation prices (such because it infrastructure) over a bigger plus base.

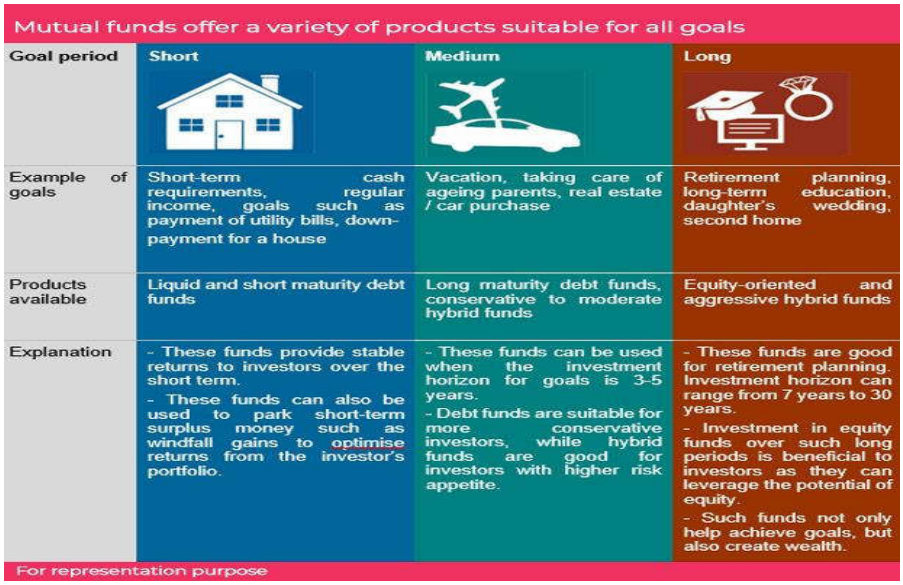
Haque, Tariq H. and Ahmed, Abdullahi D (2015) Found that Australian open-end investment company investors ought to avoid high fee funds as these funds generate relatively low after-fee risk-adjusted returns every unconditionally and in weak economic conditions.High-fee living Australian wholesale funds perform comparatively powerfully in each weak

economic conditions and flatly. High-fee funds in different styles of Australian Mutual

Funds usually don't perform powerfully either in weak economic conditions or categorically. Among low-fee funds, it had been typically true that those that perform poorly categorically but well in weak economic conditions can charge over those that perform poorly categorically and poorly in weak economic conditions.

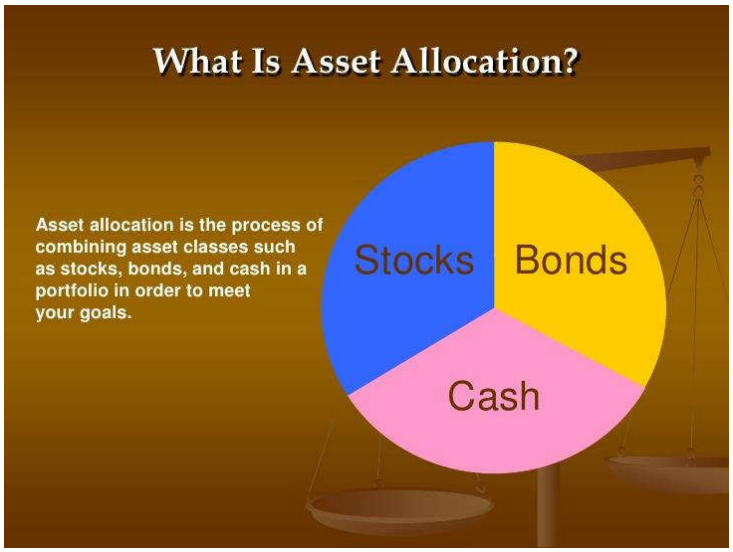
Patwa, Prerna and Agarwal, Kshama (2014) Evaluated the performance of equity funds by analyzing a sample of 4 firms every from each the sectors and 5 schemes of comparable nature. It evaluated the risk-return profile of the funds. Testing the hypotheses victimization Mann-Whitney U-test, the study discovered that there is a important distinction between the performances of personal and public sector Mutual Funds and that the personal sector has performed higher than the general public sector.

Agrawal D (2011) In his paper titled live performance of Indian Mutual Funds, delineate regarding Mutual Funds size and plus allocation. The paper analyzes Indian open-end investment company business valuation mechanism with empirical studies on its valuation. It to boot analyzes data at every the fund-manager and fund-investor levels. The study discovered that the performance is stricken by the saving and investment habits of the folks and also the second facet the confidence and loyalty of the fund manager and rewards affects the performance of the MUTUAL FUNDS business in country.



VII. ASSET ALLOCATION:

Asset allocation funds are mutual funds that invest in a varied class of assets. These assets can be equity-oriented, debt-oriented or even other asset classes like gold, other metals, and commodities.



Asset allocation funds are meant for risk-averse investors. Equity is an important asset class to have because it helps beat inflation. But equity can also be volatile, which makes it unsuitable for all investors. This is where asset allocation funds come into the picture.

Different asset allocation funds invest a part of their assets in equities and the rest in other assets. This helps them generate good returns while reducing risks.

VIII. DIFFERENT TYPES OF ASSET ALLOCATION FUNDS

Asset allocation funds can be broadly classified into two categories:

a. Dynamic Asset Allocation Funds

These funds keep changing/adjusting the proportion of assets in their portfolio according to the market fluctuations. When one particular asset class is expected to do well, the fund increases allocation to that asset and vice versa.

b. Static Asset Allocation Funds

Static asset allocation funds have a pre-decided percentage of funds allocated to different asset classes. The most popular type of funds here are balanced funds that invest at least 65% of their assets in equities and the rest primarily in debt.

IX. IMPORTANCE OF ASSET ALLOCATION FUNDS

Investors should consider asset allocation funds because of the following reasons:

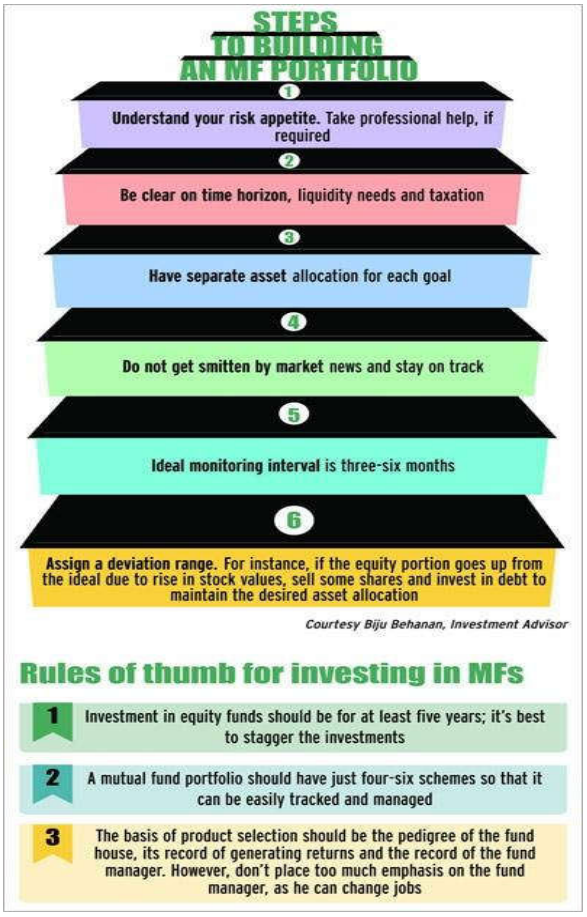
a. Diversification

Investors who want to minimize their risk for a given time frame can invest in different classes of assets and diversify their portfolio.

b. Control over volatility

If an asset class performs outstandingly in one year, there is no certainty that it will perform equally well in the next year as well. The asset class can underperform as well. By investing in a diversified portfolio, your investments will not get affected greatly by one asset class.

.



X. CONCLUSION: Asset allocation in practice, you have learnt how to determine different investor objectives with respect to risk and return and investor constraints such as investing in sustainable sectors or being concerned with taxation or currency issues. Mutual fund business is one in all the quickest growing business in Asian nation and it's already established in foreign countries. investment in Mutual Funds could be a ton of safe as compared to equity moreover because it provide handsome returns. supported differing kinds of equity and debt mutual funds, the subsequent conclusion is created.

REFERENCES:

- *"Open-end Mutual Funds and Capital Gains Taxes," Journal of monetary social science, 49, 3-43.*
- ***Author: Barclay, M., N. Pearson, and M. Weisbach Published: 1998***
- *"Conditional Market temporal order with Benchmark Investors," Journal of monetary social science, 52,119-148*
- ***Author: Becker C., W. Ferson, D. Myers, and M. Schill***

LEVELS OF MANAGEMENT IN ORGANIZATIONAL STRUCTURE

1) M.BALADIVYA
Student, Department of Business Administrations,
MRCE – MBA, Hyderabad.
E mail id : divya.muleti@gmail.com

2) D.DHARAMSINGH
Student, Department of Business Administrations,
MRCE – MBA, Hyderabad.
E mail id :d.dharamsingh@gmail.com

ABSTRACT

This title is exploring the levels of management in organizational structure. Management in all business and organization activities is the act of coordinating the effort of people to accomplish desire goals and objectives using available resources efficiently and effectively management comprises planning, organization ,staffing, leading or directing and controlling and organization. People in all of the organization operate and work in dynamic environment. But they don't operate same level they operate different level, The hierarchy of the operation in the organization is called as level's of management. A part of an organization that maintains responsibility for the productivity and the work performance of employees. There are generally three levels of management within an organization including top-level, middle-level, and first level that are tiered in numbers with more first level managers, a smaller amount of middle managers, and less top-level managers within one organization. Each level possesses certain job responsibilities within their position to ensure the effective overall operation of the organization.

I.OBJECTIVE: The objectives of this lesson are to enable to define management; to describe the nature and scope of management; to know the difference between management and administration; to understand various levels of management; and to describe the various skills that are necessary for successful managers.

II.INTRODUCTION

The term LEVLES OF MANAGEMENT refers to a line of demarcation between various managerial positions in an organization. The number of levels in management increases when the size of the business and force increases and vice versa. The levels of management determines a chain of command, the amount of authority and status enjoyed by any managerial position. In an organization three levels of management usually identified. Management is essential not only for business concerns but also for banks, schools, colleges, hospitals, hotels, religious bodies, charitable trusts etc. Every business unit has some objectives of its own. These objectives can be achieved with the coordinated efforts of several personnel. The work of a number of persons are properly co-ordinate to achieve the objectives through the process of management is not a

matter of pressing a button, pulling a lever, issuing orders, scanning profit and loss statements, promulgating rules and regulations. Rather it is the power to determine what shall happen to the personalities and happiness of entire people, the power to shape the destiny of a nation and of all the nations which make up the world. Management is a vital aspect of the economic life of man, which is an organized group activity. It is considered as the indispensable institution in the modern social organization marked by scientific thought and technological innovations. One or the other form of management is essential wherever human efforts are to be undertaken collectively to satisfy wants through some productive activity, occupation or profession.

III. DEFINITION OF MANAGEMENT

although management as a discipline is more than 80 years old, there is no common agreement among its experts and practitioners about its precise definition. In fact, this is so in case of all social sciences like psychology, 5 sociology, anthropology, economics, political science etc. As a result of unprecedented and breath-taking technological developments, business organizations have grown in size and complexity, causing consequential changes in the practice of management. Changes in management styles and practices have led to changes in management thought. Moreover, management being interdisciplinary in nature has undergone changes because of the developments in behavioural sciences, quantitative techniques, engineering and technology, etc. Since it deals with the production and distribution of goods and services, dynamism of its environments such as social, cultural and religious values, consumers' tastes and preferences, education and information explosion, democratization of governments, etc., have also led to changes in its theory and practice. Yet, a definition of management is necessary for its teaching and research, and also for improvement in its practice. Many management experts have tried to define management. But, no definition of management has been universally accepted. Let us discuss some of the leading definitions of management:

Peter F. Drucker defines, "management is an organ; organs can be described and defined only through their functions".

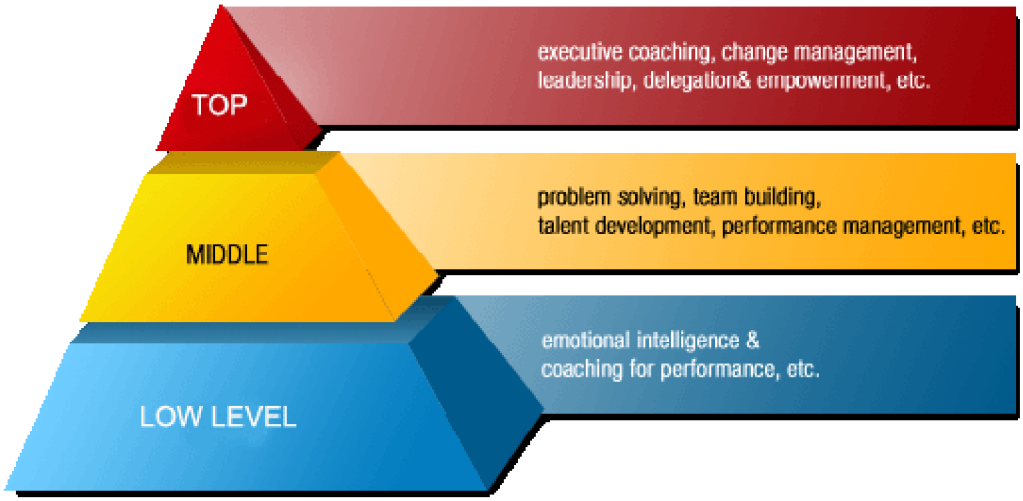
According to Terry, "Management is not people; it is an activity like walking, reading, swimming or running. People who perform Management can be designated as members, members of Management or executive leaders."

Ralph C. Davis has defined Management as, "Management is the function of executive leadership anywhere."

James Lundy, "Management is principally a task of planning, coordinating, motivating and controlling the efforts of other towards a specific objective. It involves the combining of the traditional factors of production land, labour, capital in an optimum manner, paying due attention, of course, to the particular goals of the organization."

IV. LEVELS OF MANAGEMENT :

- Top level management/Administrative level
- Middle level management
- Lower level management/ supervisory



TOP LEVEL MANAGEMENT

Top level management is a team consisting of manager from different functional levels. They are analyze, evaluate and deal with external environment. Top level management have a maximum authority in the organization and they are responsible for welfare and survival of the organization.

Roles and responsibilities of Top level management

- Top level management lays down the objectives and broad policies of enterprise.
- It issues necessary instruction for preparation of department budgets, procedure, schedules etc.
- It prepares strategic plan and policies for the enterprise.
- It controls and coordinates the activities of all the departments.
- It provides guidance and direction.

MIDDLE LEVEL MANAGEMENT

Middle level management is a second level of management. It consists of all the branch managers and departmental managers. They are responsible to the top management for the functioning of their department. They devote more time to organizational and directional functions. In small organization, there is only one layer of middle level of management but in big enterprises, there may be senior and junior middle level management

Roles and responsibilities of middle level management

- They execute the plans prepared by top level managers.

- They make plans for the sub-units of the organization.
- They participate in employment & training of lower level management.
- They give the report to the top level management.
- They diagnose and resolve problem within and among work group.
- They design and implement effective group and intergroup work and information system.

LOWER LEVEL MANGEMENT

Lower level management is third level management and it is also called as a supervisory/operative level of management. It consists of supervisors, foremen, section officers, superintendent etc. It is concern with the direction and controlling function of the management.

Roles and responsibilities of lower level management.

- They assign work to workers and they guide and instruct the workers.
- They also responsible for the proper quality and quantity of production.
- They arrange the materials, machines, tools etc.
- They make performance report of the workers.
- They spend more time in directing and controlling.
- The lower level managers make daily, weekly and monthly plans.

V. ADMINISTRATION IS A HIGHER LEVEL FUNCTION : Oliver Shelden subscribed to the first viewpoint. According to him, "Administration is concerned with the determination of corporate policy, the coordination of finance, production and distribution, the settlement of the compass of the organization and the ultimate control of the executive. Management proper is concerned with the execution of policy within the limits set up by administration and the employment of the organization in the particular objects before it... Administration determines the organization; management uses it. Administration defines the goals; management strives towards it". Administration refers to policy-making whereas management refers to execution of policies laid down by administration. This view is held by Tead, Spriegel and Walter. Administration is the phase of business enterprise that concerns itself with the overall determination of institutional objectives and the policies unnecessary to be followed in achieving those objectives. Administration is a determinative function; on the other hand, management is an executive function which is primarily concerned with carrying out of the broad policies laid down by the administration.

VI. ADMINISTRATION VS. MANAGEMENT Basis Administration Management

1. Meaning Administration is concerned with Management means getting the formulation of objectives, plans the work done through and and policies of the organization with others.

2. Nature of Administration relates to the decision- Management refers to execution work making. It is a thinking function. of decisions. It is a doing function.

3. Decision Administration determines what is to Management decides who shall Making be done and when it is to be done implement the administrative decisions.

4. Status Administration refers to higher levels Management is relevant at lower of management levels in the organization.

VII. CONCLUSION

These are the three levels in an organization top level, middle level and lower level. Top level is the highest level which has the maximum authority. Middle level is a link between the top level and lower level, Middle level receives the order from the top level and gives the instructions to the lower level and lower level consists the supervisors, foremen etc. who are in direct contact with the workers and responsible for controlling and maintaining the production. environment of an organization as well as in the outlook of the organizational members. Each member begins to see himself as a resource to others and is willing to lend his support to his colleagues when it is neededThe change must be for the better so that it develops the organization and such organizational development should be directed towards improvement in organizational health and welfare of the members. The basic purpose of such development is to improve the operational as well as interpersonal skills of employees and it also helps members to develop interpersonal competence including communication skills and an insight into themselves and others. It improves communication channels at all levels of the hierarchy and build team spirit among workers as that their inter-group relationships are highly cordial.

REFERENCE BY

1. *Peter F. Drucker has stated in his famous book "The Practice of Management" that, "the emergence of management as an essential*
2. *Louis A. Allen, Management and Organization, McGraw-Hill Kogakusha, Ltd.*
3. *Jit S. Chandan, Organizational Behaviour, Vikas Publishing House.*
4. *Fred Luthans, Organizational Behaviour, McGraw-Hill.*
5. *Stephens P. Robbins," Organizational Behaviour", Prentice-Hall India.*
6. *Laurie J. Mullins," Management and Organizational Behaviour", Pitman.*

BUSINESS ENVIRONMENT & CULTURAL DIMENSIONS IN ORGANIZATIONAL PERFORMANCES

G.PRASHANTH
Student, Department of Business Administrations
MRCE – MBA,
Hyderabad.
Mail Id : gajjelaprashanth98@gmail.com

K. SRI VENKATESH
Student, Department of Business Administrations
MRCE – MBA,
Hyderabad.
Mail Id: kanuri.venky@gmail.com

ABSTRACT: The study obverses that estimating the performance of the organization has always been of interest to management teams and researchers. In this regard, some researchers focused on determining definitions and how to measure organizational performance. The main objective of this study is to overview of organizational performance index definition and performance measurement. The history of performance is classified into six different subcategories. Various ideas and opinions on each of the performance subcategories are discussed. This review paper is applicant for researcher and student to better to understand definition and modelling of performance in organizational studies.

II. KEY WORDS: Business Environment, Organizational, Performance and Effects of Business Environment on OrganizationalPerformance, organizational performance index, organizational theory, performance measurement, efficiency, effectiveness.

III. INTRODUTION: Organizational performance comprises the actual output or results of an organization as measured against its intended outputs (or goals and objectives).According to Richard et al. (2009) organizational performance encompasses three specific areas of firm outcomes: financial performance (profits, return on assets, return on investment, etc.)Product market performance (sales, market share, etc.); and shareholder return (total shareholder return, economic value added, etc.). The term Organizational effectiveness[2] is broader.Specialists in many fields are concerned with organizational performance including strategic planners, operations, finance, legal, and organizational development. The cultural dimension of globalization' explores the intensification and expansion of cultural flows across the globe. Critics of cultural globalization claim that the world is being homogenized or 'Americanized'. Steger compares the current study of globalization to the ancient Buddhist parable of blind scholars and their first encounter with an elephant. Similar to the blind scholars, some globalization scholars are too focused on compacting globalization into a singular process and clashes over "which aspect of social life constitutes its primary domain" prevail.

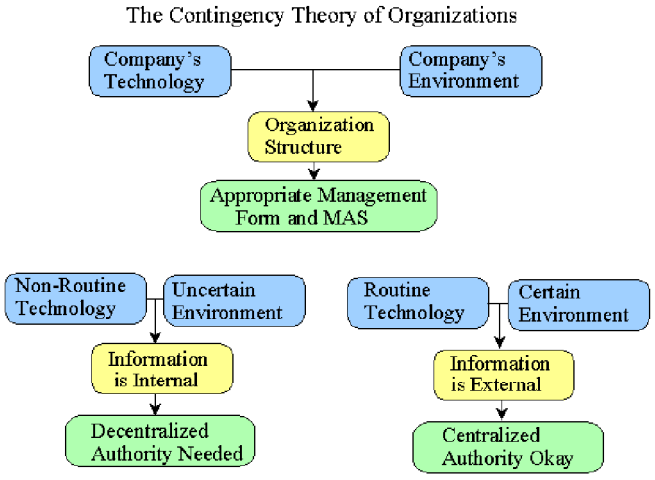
III. OBJECTIVE : One of the most critical assets of an organization used to be

machinery, but in the 21st century, it is knowledge. This knowledge can refer to creativity, ideas, expertise or experience. Employees are often the owners of this knowledge. Organizational culture can promote or hinder knowledge sharing. Multicultural organizations often face challenges in creating organizational cultures where employees feel safe to share knowledge. Organizational culture within a multicultural organization is critical to performance efficiency.

IV. ORGANIZATIONAL CULTURE'S IMPACT ON PERFORMANCE EFFICIENCY

Clear and open communication makes for an effective organizational culture. Multicultural organizations need to recognize and respect cultural rules. If cultural rules are broken, it can delay and hinder the knowledge sharing needed for performance efficiency. Being able to share the problems that prevent optimal performance and the consistency to achieve that performance can be hindered by ignoring cultural cues. By understanding and respecting cultural rules, effective leaders can create an environment where knowledge is shared freely.

V. Organizational environment is a set of forces that surrounding an **organization** that have the potential to affect the way it operates and access to scarce resources. ... Contingency theory posits that there is no one best way of structure and manage organizations.



Types of Environment

Micro Environment: the skill and ability of employees, their attitude to work, relations between managers and subordinates etc. may be regarded as internal environment of business.

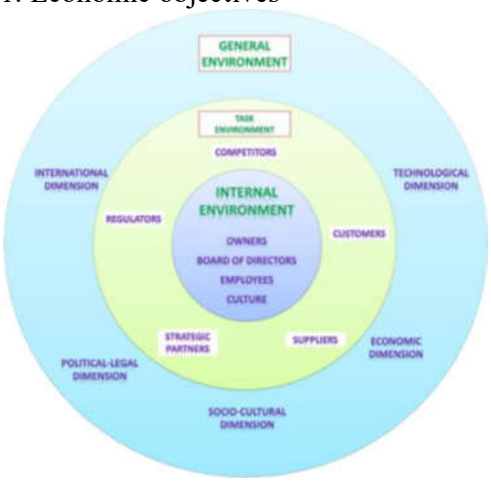
These are important factors effect business operations.

Macro Environment: on the other hand, external environment refers to all those aspect of the surroundings of business, which are not with in control of the

managers and mayaffect activities to a great extent.

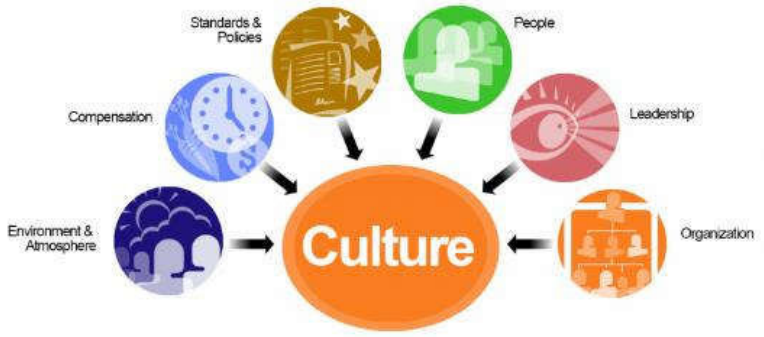
Objectives of business environment

1. Economic objectives
2. Social objectives
3. Human objectives
4. National objectives



VI. CULTURE:

A culture is a way of life of a group of people-the behaviour's beliefs, values ,and symbols that they accept, generally without thinking about them, and that are passed along by communication and imitation from one generation to the next. Culture is a symbolic communication.



Dimensions of culture:

*Neutral Vs emotional

* Neutral: culture in which emotions not shown -High neutral countries ,people act stoically and maintain composure(e.g., Japan, and u.k)

Emotional: Emotions are expressed openly and naturally

-high emotions and cultural: people smile a lot, loudly, greet each other with enthusiasm (e.g., Mexico , Netherlands ,Switzerland)

*Specific Vs diffuse

- specific: large people space shared with others and small private space guarded closely

* high specific culture: people open, extroverted; strong separation work and personal life (e.g.,Austria,U.K.,U.S)

- Diffuse: public and private spaces similar size, public space guarded because shared with private space; people indirect introverted ,work/private life closely linked (e.g., Venezuela,china,spain)

* Achievement Vs Ascription

- Achievement culture ;status based on how well perform functions (Austria,Switzerland,U.S.)

-Ascription culture: status based on who or what person is (e.g., Venezuela, China, Indonesia)

-TIME:

-Sequential: only one activity at a time: appointments kept strictly, follow plans as laid out (U.S.)

- Synchronous: multi-task, appointments are approximate, schedules subordinate to relationships (e.g., France, Mexico)

* THE Environment

- Inner-directed: people believe in control of outcomes (U.S., Switzerland, Greece, Japan)

- Outer-directed: people believe in letting things take own course(China, many other Asian countries)

VII. RECOMMENDATIONS:

- Embrace transparency. Transparency isn't just positive for employees.
- Recognize and reward valuable contributions.
- Cultivate strong coworker relationships. ...
- Embrace and inspire employee autonomy. ...
- Practice flexibility. ...
- Communicate purpose and passion. ...
- Promote a team atmosphere. ...
- Give and solicit regular feedback.

VIII. CONCLUSION:

Business environment involve internal and external environment. Business environment is important for an organization to identify the weaknesses and threats. This helps an organization to reduce the risk of getting failure in their operation and development in new product. The culture within an organization is very important, playing a large role in whether it is a happy and healthy environment in which to work. In communicating and promoting the organizational ethos to employees, their acknowledgement and acceptance of it can influence their work behaviour and attitudes.

References:

1) *Organizational Behaviour: Text and Cases – AK Chitale and RP Mohanty*
2) *Essentials of Organizational Behaviour – Pearson*

3) *Organizational Behaviour – Stephen P. Robbins and Tomothy A Judge*
4) *Cross Culture and Management – Concepts and cases – Shobhana Madhavan*
5) *A Cross culture theory of Voter Behaviour - Bruce I. Newman*
6) *Advancing Cross Culture Perspective on Educational Philosophy – George Arief D. Liem and Allam B L . Bernardo*

OVERVIEW OF E-COMMERCE AND PRESENT GLOBAL TRENDS

1) RATHOD SWAPNA LATHA
Student, Department of Business Administrations,
MRCE –MBA, Hyderabad.
E –mail: swapnarathod198@gmail.com

2) S.SHARUN
Student, Department of Business Administrations,
MRCE –MBA, Hyderabad.
E –mail:sharonsabbithi@gmail.com

ABSTRACT: E- commerce stands for electronic commerce. In other word sit is called as e-trade or e- business. E- commerce is a transaction of goods and services conducted electronically between business and customer over computer network such as internet. This saves time for participants on both ends. It has a pretty good role in our daily life also. In daily life there are many applications of e- commerce. The main services provided by e-commerce are e- mail, video- conferencing, electronic shopping, electronic banking etc. We can see how important it is for any kind of business. In this paper we will discuss about how e-commerce effects the consumers and business both positively and negatively, and we also talk about the types of e-commerce. More consumer demands appear in the internet than any other transmission mediums today. According to Morgan Stanley U.S. Investment Research, the internet only takes four years to reach fifty million people while television take more than ten years and radio about 38 years. It demonstrates that the internet has a huge potential to grab many customers in a short amount of time, which in turn can become a target market for growing firms.

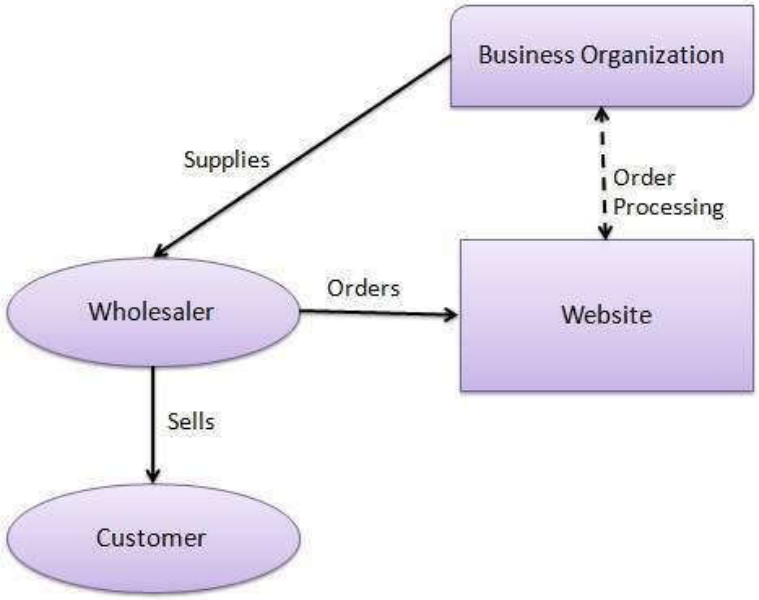
I.INTRODUCTION: Electronic commerce is all started in 1979 by MICHAEL ALDRICH an english inventor, innovator and entrepreneur. He enabled online transactions by inventing online shopping. It is the process of execution of commercial transactions electronically with the help of technologies of EDI electronic data interchange, EFT electronic fund transfer, supply chain management, internet marketing etc….In these cases customer deals directly with producer/sellers and intermediaries who offers a wide range of goods and services.E-commerce is the largest sector of the electronics industry. It is generally used to improve the efficiency and effectiveness of a company's sales efforts. Instead of taking orders from customers manually like by making telephone calls or meeting personally, orders are received digitally through internet.

II. KEY WORDS: Types of e-commerce, pros and cons of e-commerce, top e-commerce websites in the world.

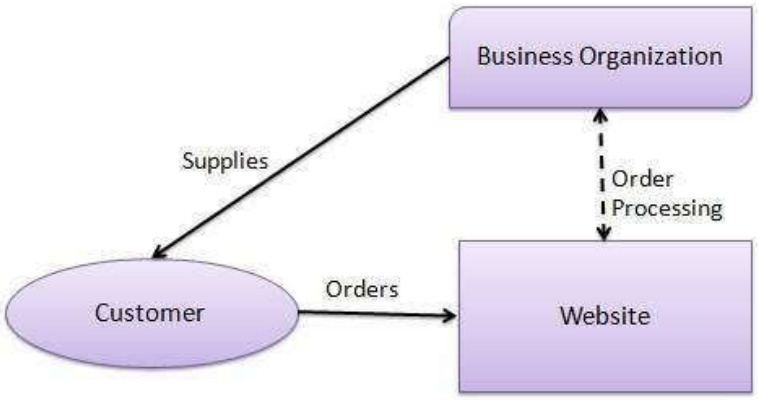
III. TYPES OF E-COMMERCE: It is basically divided into six types

- Business To Business
- Business To Customer
- Customer To Customer
- Customer To Business
- Business To Government
- Government To Business.

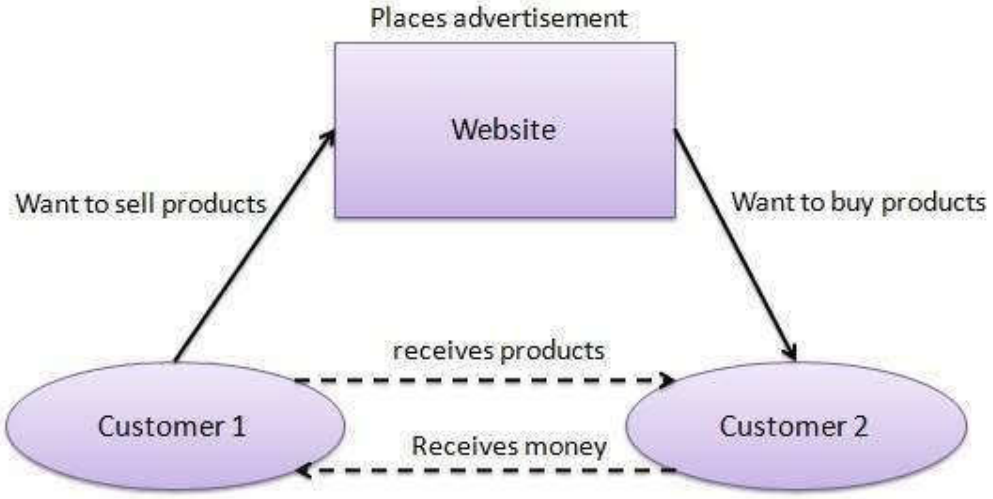
BUSINESS TO BUSINESS(B2B):In this case both the seller and the buyer are business entity. In this the business organisation will sell its product to another business organisation/intermediate who will sell goods to the final consumer B2B covers a large number of applications, which enables business to form relationships with their distributors, re-sellers, suppliers, etc. Examples for B2B are bulk bookstore, kick starter etc..



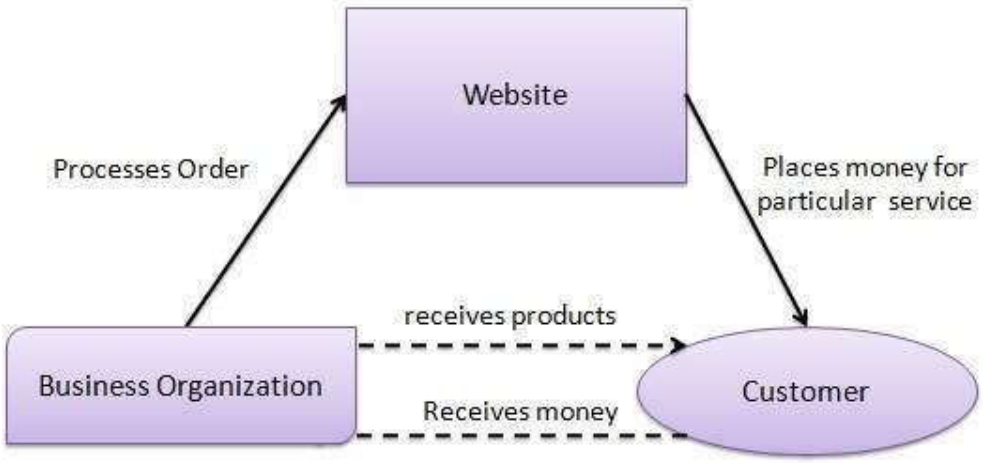
BUSINESS TO CUSTOMER(B2C):In this case,the seller is a business organisation whereas the buyer is the customer. Typically electronic stores are set up on the internet to sell goods to customers. Some of the examples of B2C are AMAZON,FLIPKART, WALMART etc….B2C not only includes the products but it also offers services as well. The decision-making process for a B2C purchase is a much shorter than a business-to-business (B2B) purchase. In business to customer the customer first visit the website of the product which he want then he will compare the prices and qualities of the product, he will select the product place the order, then the organisation will accept the order and dispatches the product.



CUSTOMER TO CUSTOMER(C2C):This refers to a situation where buyer and seller are customers. with the help of e-commerce, online auctions provide an effective means for supporting C2C e-commerce. In this model the customer can sell their property, car, motor vehicles etc..Or rent a room by providing information on the website. Some websites will not charge for its service. some other customer may opt the product of the first customer by seeing the advertisement on the website. Examples are OLX,QUICKER and other websites.



CUSTOMER TO BUSINESS(C2B):this is a new form of commerce in which a consumer specifies the requirements to a business, which provides a product that means these requirements. these requirements could be as simple as an acceptable price, customisation of an existing product or creating a new product. This type of E-commerce is mostly used for individual consumer giving services or products to an organization or a company. Especially if you from software field, you can give your services and use your skills by using some website like freelancer, fiver etc, the companies are waiting to hire a skilled service and all this is done by C2B model.



IV. PROS AND CONS OF E-COMMERCE:

PROS TO ORGANISATION:

1. Organisation can expand their market globally with less capital investment and it can easily locate more customers, suppliers and business partners globally.
2. E-commerce helps in reducing the paper based information bt digitalsing the info.
3. E-commerce provides better customer service and also improves brand image of the company.

PROS TO CUSTOMERS:

1. Customer can buy the product or service at anytime from anywhere 24*7.
2. E-commerce provide many option with quick delivery to the customer.
3. It provides many option customer can compare and select cheaper and better quality.

PROS TO SOCIETY:

1. Customer need not travel to buy a product, thus reduces traffic and air pollution.
2. E-commerce helps in reducing the cost of production, so everyone can effort the product.
3. E-commerce enable rural areas to access services and products.

CONS OF E-COMMERCE:

1. SECURITY: Many people fear of providing information regarding personal and financial even though lot of improvements are made regarding securing customer data Certain websites do not have capabilities to conduct authentic transactions.

2. TAX ISSUES**:** Sales tax is another bigger issue when the buyer and seller are in different locations the computation of sales tax poses problems.

3. PRODUCT SUITABILITY: People have to believe on the images shown, and have to purchase the product. Sometimes the product will be different from the image shown on internet. LACT OF TOUCH AND FEEL keeps the people far from online shopping.

4. HIGH LABOUR COST :To develop and manage the website of the organisation highly talented and technical work force is required. The organisations have to incur lot of expenses to retain talented people.

5. SEVERE COMPETITION : There are unimaginable portals for a single product because of that the competition increases .this forces the company to reduce their prices by discounts.

V. TOP E-COMMERCE WEBSITES IN THE WORLD:

1. AMAZON
2. WALMART
3. EBAY
4. TARGET
5. ALIBABA
6. FLIPKART
7. NEWEGG

VI. CONCLUSION:

E-commerce is continuously progressing and is becoming more and more important to **businesses** as technology continues to advance and is something that should be taken advantage of and implemented. From the inception of the Internet and **e-commerce**, the possibilities have become endless for both **businesses** and consumers. However, there are some negatives about e-commerce. Despite security protocols (like HTTPS); businesses with the internet were more likely to be the victims of security loss. Information loss, computer virus, theft/hacking, and manipulating their internal systems, are big issues of today's e-commerce. They try to secure all transactions with SSL (Secure Sockets Layer for private transactions through internet) security standards. Hosting your store in a secure hosting environment or displaying a disclaimer during the checkout process helps to ensure that the visitor is in a safe environment. However, intellectual property protection still seems to be an issue of today. E-Commerce is an evolution that helped our traditional economy adapt to the new digital world, using electronic technology through the internet, resulting in more competitions, more marketplaces, faster transactions, and more advanced technologies to make activities between customers and producers more active. Internet has changed economy so much that most of the business activities today are made in the web. Therefore, we as customers and internet users are responsible to keep our e-commerce healthy and safe so that e-business can be more reliable in the

future. That way, every single person in the world will be able to take advantage of e-commerce.

REFERENCE:

Ecommerce Issues. GoECart.com. 3 Nov 2007. <http://www.goecart.com/ecommerce-issues.asp>.

Electronic Commerce. 8 Nov 2007. Wikipedia. 10 Nov 2007. <http://en.wikipedia.org/wiki/Electronic_commerce>.

E Commerce evolved: The Essentials by Tanner Larsson

E Business and E Commerce Management by Dave Chaffey

You Should Test That: Conversion Optimization for More Leads, Sales and Profit Or The Art and Science of Optimized Marketing by by Chris Goward

Study on Importance of Soft Skills, A pathway for successful career

POOJA SINGH
Assistant Professor,
Malla Reddy College of Engineering, MRCE, Maisammaguda, Hyderabad 500100
e-Mail: pooja.amitsingh@gmail.com

ABSTRACT:

This paper makes a survey of the importance of soft skills in students' lives both at college and after college. It discusses how soft skills complement hard skills, which are the technical requirements of a job the student is trained to do. The paper exhorts educators to take special responsibility regarding soft skills, because during students' university time, educators have major influence on the development of their students' soft skills. Embedding the training of soft skills into hard skills courses is a very effective and efficient method of achieving both an attractive way of teaching a particular content and an enhancement of soft skills. Soft skills fulfil an important role in shaping an individual's personality. It is of high importance for every student to acquire adequate skills beyond academic or technical knowledge.

I.INTRODUCTION:

For decades employers as well as educators frequently complain about lack insoft skills among graduates from tertiary education institutions.

Research conducted by Harvard University, the Carnegie Foundation and Stanford Research Centre have all concluded that 85 per cent of job success comes from having well-developed soft and people skills, and only 15 per cent of job success comes from technical skills and knowledge (hard skills).

These statistics were extrapolated from a study published in 1918 by the Carnegie Foundation. For almost 100 years, we learned that soft skills are critical to the success of any organization.

Based on a number of surveys on the skills required by graduates undertaken by Microsoft, Target Jobs, the BBC, Prospects, NACE and AGR and other organisations, the skills which are deemed most important are:

- Ability to express ideas clearly and confidently in speech
- Work confidently within a group
- Understand the commercial realities affecting the organisation
- Gather information systematically to establish facts and principles for problem solving
- Ability to take initiative, identify opportunities and being proactive in putting forward ideas and solutions
- Determination to get things done, make things happen and constantly look for better ways of doing things
- Ability to express thoughts and ideas clearly in writing
- Ability to plan and execute effectively
- Adapt successfully to changing situations and environments
- Manage time effectively, prioritise tasks and work to deadlines

II. WHAT ARE SOFT SKILLS?

Before going any further in debating the importance of soft skills we have to clarify the question "What exactly are soft skills?" This basic question is not easy to answer, because the perception of what is a soft skill differs from context to context. A subject may be considered a soft skill in one particular area, and may be considered a hard skill in another. On top of it the understanding of what should be recognised as a softskill varies widely. Knowledge in project management for instance is "nice to have"for an electrical engineer, but it is a "must to have" for a civil engineer. Training incultural awareness might be useful for a chemist, but it is an absolute necessity forpublic or human resources management in societies of diverse cultures.

Interesting enough the internationally renowned encyclopaedias have little to sayabout soft skills. The online encyclopaedia "Wikipedia" gives a very broad definitionof soft skills, which leaves much room for discussion:

"Soft skills refer to the cluster of personality traits, social graces, facility withlanguage, personal habits, friendliness, and optimism that mark people to varyingdegrees. Soft skills complement hard skills, which are the technical requirementsof a job." (Wikipedia, 2007)

Soft skills are in two types.

Internal Soft Skills:

- Self-confidence
- Self-awareness
- Self-compassion
- Accepting criticism
- Critical thinking/problem solving
- Resilience
- Perseverance
- Emotional management
- Perceptiveness
- Growth mindset
- External Soft Skills

External Soft Skills:

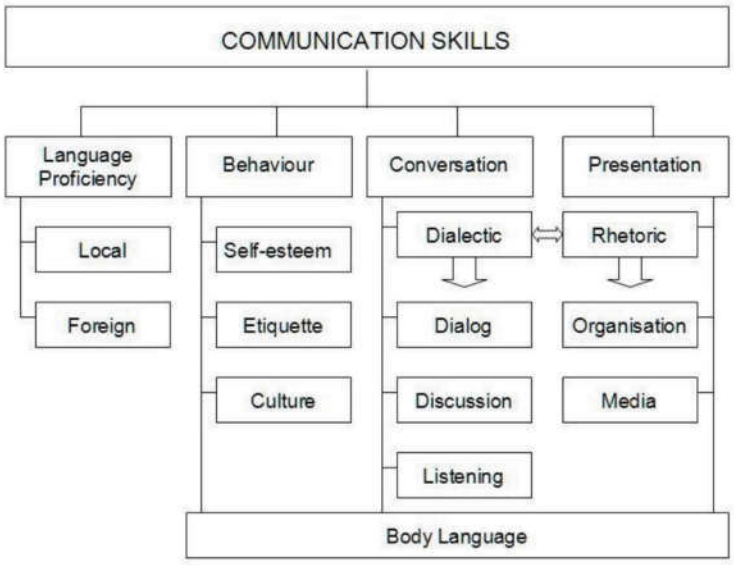
- Collaborative teamwork
- Effective communication
- Interpersonal skills
- Self-Promotion
- Managing conflict
- Adaptability
- Networking
- Influence
- Negotiation
- Expectation management

Above list of examples of soft skills based on the Wikipedia definition.

What are the most important soft skills?

Again the answer regarding the importance of soft skills depends very much onthe context and one's personal perception. However, there is one property thatimmediately comes to most people's mind when soft skills are mentioned: thoseare the communication skills. And indeed, it is the talent of communication skills, which is mostly lacking among graduates from colleges and universities.

When asking people what exactly they understand to be communication skills,one will receive a wide range of answers, because communication skills include a lot of different aspects. Following figure offers an overview of important factors shaping a person's communication skills.



The collection of aspects spans from basic language proficiency, which in multi-lingual societies like Namibia may not be taken for granted, to advanced topics likeDialectic or Rhetoric, which are sophisticated skills in their own right. Minimumrequirements for a graduate of a tertiary institution should be ample proficiencyinspoken and written language, a certain amount of self-esteem that will be reflected in conversation skills and body language, adequate discussion skills,and of major importance, good presentation skills in order to be able to marketoneself and one's ideas. However, communication skills are not only necessary for aperson's professional career, but are even more contributing to one's so-called socialcompetence, a fact which applies to many other soft skills, too. Good social skills arealso reflected at the working place and hence recursively further the career.However, if we continue drilling down deeper into Communication skills theissue will become even more complex. In the English speaking world "Languageproficiency is the ability to speak, read, and write Standard English in a business like
way. One may have the 'hard' skill of knowing what usage is correct and what isincorrect, but lack the 'soft' skills of knowing when to use only standard forms andin what tone to use them." (Waggoner, 2002). In this definition our basic soft skill"Communication Skills" itself is divided into a "soft" and a "hard" part.
Furthermore, adequate communication skills are a prerequisite for a range of othersoft skills like moderating discussions or conflict management.

Another pair of soft skills frequently lacking in tertiary education is critical andstructured thinking. Both go hand in hand with problem solving abilities. Especiallyin today's information society it is of high importance to critically filter the endlessstream of incoming data, analyse it, and make informed decisions based on it.Analytical skills also form the base for developing solutions to any kind of problem.And also in this case the soft skills are of equal usefulness in a person's professionaland private life.

A last soft skill that should be highlighted here is creativity. This skill is oftenmisinterpreted as being only useful for artists, whereas in the science or businessarena only structured logical thinking should be applied. However, this perceptionis wrong. Applying creativity results in "thinking out of the box", which means thatgiven conventional rules and restrictions are left aside in order to find innovativeapproaches to problem solving. If Albert Einstein had not bypassed the guidelinesof Newton's old established physics, and even discarded a bit of common sense,he probably would not have arrived at his revolutionary new view on physics.Brainstorming and mind mapping are well known applications of creativity in the
business world.

III. CAN PERSONAL TRAITS AND HABITS BE CHANGED?

As we recognised earlier, a large part of soft skills relates to personal traits andhabits. Thus, an interesting question to ask is whether a person's traits and habitscan be changed or improved. Anybody raising children or living with a partner whohas irritating habits, might be tempted to answer that the task of changing personaltraits is extremely difficult, if not impossible. The corporate trainer and instructionaldesigner Rukmini Iyer has a strikingly simple answer to this question: "There is alot of argument in industry as to whether it is possible to enhance soft skills in afew hours of training, especially when one considers the fact that a person has livedwith those traits all his life. To this, the answer is harsh but real -- a professional whowants to do well in his / her career does not really have a choice." (Iyer, 2005)

A key aspect in this regard is a person's ability to recognise and acknowledgecertain behavioural shortfalls or plain bad habits. As a German proverb states:"**Self-recognition is the first step towards improvement**". As soon as a deficit hasbeen identified, one can start working on its elimination.

For example, many people are not fond of doing small talk, a common characteristic among those working in a scientific or technical environment. The reasons might vary shyness, introverted character, lack of "how-to", boredom, etc. However, small talk is an important part of the communication skills necessary to display social competence. Once having acknowledged this truth, a person can undergo a simple self-training or guided training to improve the lacking skill. There are for instance many books available on the issue of "Small Talk". Forthwith, the person can make it a habit at parties or social gatherings to initiate a conversation with at least three strangers. While the first times might require quite an effort, after a few months the frequent chatting with strangers will become second nature. With such and similar easy exercises even a very introverted person can appear reasonably sociable.

In summary, it means that negatively perceived personal traits could be changed or successfully covered
by undergoing self-imposed training. The only prerequisite is that one acknowledges one's weakness and takes the decision to change it. Training will most likely be unsuccessful if one is not fully convinced

that it will lead to any improvement or that the improvement will be beneficial.

IV. WHY ARE SOFT SKILLS IMPORTANT?

After having elaborated so much on soft skills, the answer to why they are consideredas being so important is still open. There are numerous reasons for having a criticallook at a person's soft skills.

One straightforward reason is today's job-market, which in many fields is becomingever increasingly competitive. To be successful in this tough environment,candidates for jobs have to bring along a "competitive edge" that distinguishesthem from other candidates with similar qualifications and comparable evaluationresults. And where do they find this competitive advantage? In bringing along
additional knowledge and skills, added up by convincing personal traits and habits.
This sounds familiar.
Understandably, employers prefer to take in job candidates who will be productivefrom a very early stage on. If a graduate from university first has to be trainedon putting more than three sentences together,how to do a proper presentation,or how to chat in a pleasant and winning manner with colleagues andcustomers,this graduate will not qualify as a quick starter. Also basic knowledge in businessmanagement, project management and general economy will improve the chancesof a job candidate considerably.

Already during the job interview itself good communication skills are invaluable.They can even serve to successfully cover up weaknesses on the hard skills side.Don't we all know colleagues who are splendid talkers, but there is no actionforthcoming from their side? The advantages of displaying positive traits likecourtesy, honesty, flexibility, common sense, flawless appearance, etc. during a jobinterview have not even to be discussed.

Vice versa, it unfortunately happens very seldom that a job candidate who wasrejected because of inadequate soft skills is told the truth about the reason forrejection, e.g. "Your body language showed that you seemed to feel very insecure,and you had problems to express yourself and present yourself in a convincingmanner". Even so this evaluation may sound harsh at first glance; this kind offeedback would help a job seeker a lot in improving his/her style for oncominginterviews. However, instead employers usually give no reason or even vaguemisleading reasons for rejections, which are of no help at all for the unluckycandidates.

Once employed, the success story of people who know how to master soft skillscontinues because of much better career opportunities. Simple fact, which can be verified in daily business life, is that employers prefer to promote staff members with superior soft skills. Good hard skills alone are not necessarily enough anymore to be a first choice when it comes to promotion. Soft skills are shaping human beings' personality. Any educator's dream is that graduates, especially from tertiary education institutions, should not only beexperts in a certain field but matured personalities with a well-balanced, roundedoff education.

However, this characteristic is reflected in soft skills, not in hardskills.

During the last decades in many societies the opinion on soft skills has changedconsiderably. Whereas in the past the mastering of hard skills was rated first andsoft skills were considered as "nice to have", the perception has been turned upsidedown. As mentioned before, good communication skills can easily be used to coverup a lack in hard skills. Nowadays in general, people who are extroverted, who aregood in marketing themselves, and who are socialising easily are rated superiorto others who lack those attributes. The good old technician, an ace in his field,but being introverted and talking less than ten complete sentences a day is notappreciated any longer.

This development is not necessarily positive, and it must be allowed to ask thequestion, whether today soft skills are overemphasised.

V. WHERE CAN SOFT SKILLS BE ACQUIRED?

We already identified two methods of learning or improving soft skills. One wayis enrolling for formal training, e.g. taking evening classes on Rhetoric, languages,presentation skills, conflict or cultural management. This is a well-proven conceptwith the advantage of having some kind of certificate at the end of the course, whichmight come in handy for job applications. The slight sarcasm in the last sentence
is intentional, because it is not always guaranteed that a certain course actuallysuccessfully enhanced a person's soft skills.

The other way of acquiring soft skills we looked at is self-training, usually based onbooks. As we saw earlier, changing of personal traits often requires long-term practiceand therefore self-training might be more useful regarding the improvement of thiscategory of soft skills. Additionally, during the last decade another method of self-training has become increasingly popular: Electronic Learning, usually abbreviatedto e-learning. The article "Yes, web based training can teach soft skills" (Horton,2007) emphasises the practicality of this approach.

A very pleasant way of self-training one's soft skills is frequent socialising withfriends, colleagues and other members of society. This may sound astonishing,but meant here is socialising consciously, i.e. with the purpose in mind to enhancecertain soft skills. We know already that in this way we can improve our small talkcapabilities, but there are a lot more soft skills especially related to Communication
skills, which can be practiced while chatting and discussing with others in aninformal manner: e.g. language proficiency in general, listening, discussing,etiquette, self-esteem, or body language.

Hopefully, at most education institutions at all levels teaching methodology hasbeen changed or will be changed towards more student centred learning. Such ashift goes hand in hand with embedding soft skills into the teaching of hard skills.

VI. WHAT CAN LECTURERS DO TO ENHANCE THEIR STUDENTS' SOFT SKILLS?

A first step in improving soft skills of students is to raise their awareness aboutthe importance of soft skills and the consequences of shortcomings in this regard.Students should be encouraged to enhance their soft skills by applying the methodswe mentioned before, e.g. reading dedicated books, attending courses, and joiningclubs or societies to broaden their horizon, like debating societies, Toast Masters
who are practicing Rhetoric, or scientific societies who offer presentations anddiscussions.

A formal approach to the problem would be to incorporate soft skills subjects intoa programme's curriculum. On lower levels a course that requires students to do abit of research and to present their results to the class afterwards has been proven asbeing quite effective. On graduate level a course on management skills, includinge.g. some communication skills together with the management of time, conflict,
cultural issues, and of major importance, oneself, has in practice been well receivedby students.

However, very often the curricula are already overloaded with hard skills courses,making it almost impossible to add or substitute courses. Furthermore, otherlecturers might be ignorant of the importance of soft skills and hence, do not supportdedicated courses in this regard. A very elegant way of offering soft skills trainingto students is to embed it into the teaching of hard skills. This way, no changes
to a programme's curriculum are necessary; instead the change will be reflectedin the lecturers' teaching methodology. An increase in group discussions, lettingstudents do presentations, and using special methodologies like De Bonoor Neuro-Linguistic Programming (NPL)can be applied throughout a course. This approachto practicing soft skills requires some re-thinking and re-planning of existing hard
skill courses. During a presentation at the Polytechnic of Namibia Professor RobKruegeroffered an interesting approach to the problem by turning the purpose ofa lecture upside down, i.e. making the content of a lecture the vehicle to teach softskills. An example would be a mathematics teacher who plans a lecture by firstlydetermining what soft skills s/he wishes to enhance on that day, and then secondly
considers how the required mathematics content can be arranged to support thisgoal. Correctly applied, such an approach to teaching will automatically increasethe attractiveness and effectiveness of a course regarding both, hard skills and softskills.

VII. CONCLUSION

Considering the fact that during the last decades in society the perceived importanceof soft skills has increased significantly, it is of high importance for everyone toacquire adequate skills beyond academic or technical knowledge. This is notparticularly difficult. Once a shortcoming in a certain area of soft skills has beenidentified at oneself, there are numerous ways of rectifying such a deficiency.Educators have a special responsibility regarding soft skills, because duringstudents' School and University time they have major impact on the developmentof their students' soft skills. Besides raising awareness regarding the importance ofsoft skills and encouraging students to improve their skills, lecturers

should activelypractice soft skills with their students. A very effective and efficient way of doingthis is to include soft skills training into the teaching of hard skills. As a positiveside effect the lessons will become more attractive, which in turn will increase thesuccess rate of learners.

Soft skills fulfil an important role in shaping an individual's personality bycomplementing his/her hard skills. However, over-emphasising it to such an extentshould not taint the importance of soft skills, that hard skills, i.e. expert knowledgein certain fields, are demoted to secondary importance.

References:

INTERNET RESOURCES

British Association of Graduate Recruiters (AGR): www.agr.org.uk

Horton W., Horton K.: www.horton.com

Iyer,Rukmini:http://us.rediff.com/getahead /2005/jun/30soft.htm

Phani,ChallaRam: http://in.rediff.com/getahead/2007/jan/08s oft.htm

PISA: www.pisa.oecd.org

Waggoner, Jaqueline: http://mountainrise.wcu.edu/html/waggone r.pdf

Wikipedia:www.en.wikipedia.om

RECRUITMENT AND SELECTION

KRISHNAVENI,
Student, Department of MBA,
Malla Reddy College of Engineering,
Hyderabad.
Mail id: svaishnavi415@gmail.com

K.B.S.S.VAISHNAVI
Student, Department of MBA,
Malla Reddy College of Engineering,
Hyderabad.
Mailid:svaishnavi415@gmail.com

ABSTRACT

Everything you need to know about the recruitment and selection process in HRM. One of the main recruitment and selection is Human resource planning ensures that right kind and right quality of employees are employed so as to achieve the organizational goals. Recruitment and selection are the two crucial steps in Human Resource process. Recruitment is an activity of establishing contact between employers and applicants. Selection is a process of picking up more competent and suitable employees.

I. KEY WORDS

Recruitment, Selection, Candidates, Organization, Jobs, Factors, Interviews

II. INTRODUCTION

The importance of ensuring the selection of the right people to join the workforce has become increasingly apparent as the emphasis on people as the prime source of competitive advantage has grown. Beaumont (1993) identifies three key issues that have increased the potential importance of the selection decision to organizations. First, demographic trends and changes in the labour market have led to a more diverse workforce, which has placed increasing pressure on the notion of fairness in selection.

Second, the desire for a multi-skilled, flexible workforce and an increased emphasis on team working has meant that selection decisions are concerned more with behavior and attitudes than with matching individuals to immediate job requirements. And third, the emphasis between corporate strategy and people management has led to the notion of strategic selection: that is, a system that links selection processes and outcomes to organizational goals and aims to match the flow of people to emerging business strategies.

Selective hiring (i.e. the use of sophisticated techniques to ensure selection of the 'right' people) is frequently included in the 'bundles' of best HR practice (see, for example, Pfeffer, 1998). The contribution of effective recruitment and selection to enhanced business performance is also illustrated by the findings of empirical studies.

For example, a study into small and medium-sized manufacturing establishments

(Patterson et al., 1997) found the acquisition and development of employee skills through the use of sophisticated selection, induction, training and appraisals to have a positive impact on company productivity and profitability. Thus the practice of recruitment and selection is increasingly important from an HRM perspective.

At the same time, however, many of the traditional methods of recruitment and selection are being challenged by the need for organizations to address the increased complexity, greater ambiguity and rapid pace of change in the contemporary environment. This chapter, therefore, discusses key contemporary approaches to recruitment and selection, and examines the influence of external and internal factors on the process. After clarifying what we mean by recruitment and selection, we begin by describing the external context in which recruitment and selection occur, including government policy and legislation.

Next, we turn our attention to the internal organizational context in order to examine factors that might account for variations in recruitment and selection practice. We then explore the systematic approach to recruitment and selection, and discuss recent developments at each stage of the process. In the final section we emphasis the two-way nature of recruitment and selection, and consider ethical issues in the treatment of individuals. The chapter concludes with a summary and a number of self-test exercises.

III. DEFINITION

The recruitment and selection process is concerned with identifying, attracting and choosing suitable people to meet an organization's human resource requirements. They are integrated activities, and 'where recruitment stops and selection begins is a moot point' (Anderson, 1994). Nevertheless, it is useful to try to differentiate between the two areas: Whitehill (1991) describes the recruitment process as a positive one, 'building a roster of potentially qualified applicants', as opposed to the 'negative' process of selection.

So a useful definition of recruitment is 'searching for and obtaining potential job candidates in sufficient numbers and quality so that the organization can select the most appropriate people to fill its job needs' (Dowling and Schuler, 1990); whereas selection is concerned more with 'predicting which candidates will make the most appropriate contribution to the organization – now and in the future' (Hackett, 1991)

IV. PRINICIPLES OF RECRUITMENT AND SELECTION

When focusing on the recruitment and selection of employees for your business, there are some general considerations you should always keep in mind:

- The first point to recognize about recruitment is that it is a process with a number of key stages, all of which combine to enhance your chances of finding the best

candidates available for any advertised position.

- It is also worth pointing out that in terms of leading and managing employees if you are not recruiting the best people available, then it is always going to be an uphill struggle to manage them day-to-day.
- Another general rule is that when seeking to fill any vacancy you should always consider the internal candidates that could be promoted to the available post and then recruit externally for the more junior position.
- Too often senior managers pay scant attention to the recruitment process and only become actively involved when a senior post is being filled, or at the end of the process for a quick 'final' interview. This is a mistake and you should be concerned with the quality and suitability of every employee who joins your business.
- It is often assumed that interviewing is something that any experienced manager can do. Again, this is a mistake: yes, anybody can conduct an interview, but few can do it well unless they are appropriately trained. Nobody in your business should conduct interviews without adequate training.
- There are many legal issues associated with the recruitment process and you should familiarize yourself with all relevant legislation.

The principle objective of the recruitment process should be to recruit, select and appoint employees appropriate to the present and future needs of your business.

V. FACTORS AFFECTING RECRUITMENT AND SELECTION:

Recruitment is a significant purpose of the Human Resource Management in a business, and it is ruled by a combination of numerous factors. Active HR Professionals must understand these factors manipulating the recruitment and take required actions for the improvement of the business.

At times when market condition changes, the organization also wants to screen these alterations and learn how it have emotional impact the incomes and analyze these functions for making recruitment an real process.

We have Internal Factors as well as External Factors that affect the recruitment procedure. In this we will be debating these factors in detail.

➢ INTERNAL FACTORS:

Organizations have control over the internal factors that affect their recruitment functions. The internal factors are −

- Size of organization
- Recruiting policy
- Image of organization
- Image of job

• SIZE OF ORGANIZATION:

The most significant issues moving the recruitment process are to expand the size of the organization, recruitment planning is

compulsory for hiring more resources, which will be handling the future operations.

- **RECRUITING POLICY:**

Signing from internal or external bases of group is also a factor in recruitment policy of an organization, which touches the recruitment process. It states the purposes of the recruitment and offers a framework for the application of recruitment programs.

- **IMAGE OF ORGANIZATION:**

Organizations having a decent positive image in the market can effortlessly entice competent resources. Keep up good public relations, given that public services, etc., certainly helps an organization in improving its standing in the market, and thereby appeal to the best likely resources.

- **IMAGE OF JOB:**

Just like the image of the institute, the image of a job plays a precarious role in employment. Jobs having a positive brand in terms of better payment, advancements, credit, and good work environment with career growth opportunities are considered to be the characteristics to entice qualified candidates.

- **EXTERNAL FACTORS:**

External factors are individuals that cannot be measured by an organization. The external factors that disturb the recruitment process include the resulting −

- **DEMOGRAPHIC FACTORS** − Demographic factors are associated to the features of possible employees such as their age, religion, literacy level, gender, occupation, economic status, etc.
- **LABOR MARKET** − Labor market panels the demand and supply of labor. For instance, if the supply of people having a precise skill is less than the demand, then the employing will need more hard work. On the other hand, if the demand is less than the supply, the hiring will be relatively easier.
- **UNEMPLOYMENT RATE** − If the unemployment rate is high in an exact area, hiring of capitals will be simpler and easier, as the number of candidates is very high. In contrast, if the unemployment rate is low, then recruiting tends to be very difficult due to less number of resources.
- **LABOR LAWS** − Labor laws replicate the social and political surroundings of a market, which are produced by the central and state governments. These laws command the compensation, working environment, security and health regulations, etc., for dissimilar types of employments. As the government changes, the laws to change.
- **LEGAL CONSIDERATIONS** − Job reservations for different castes such as STs, SCs, OBCs are the best instances of legal concerns. These concerns, passed by government, will have a positive or negative impact on the recruitment policies of the organizations.

- **COMPETITORS** – When governments in the similar business are contending for the best capable resources, there is a requisite to examine the competition and offer the resources packages that are best in terms of industry values.

VI. RECRUITMENT AND SELECTION PROCESS

The five steps involved in recruitment process are as follows:

(i) Recruitment planning (ii) Strategy Development (iii) Searching (iv) Screening (v) Evaluation and Control.

TYPES OF RECRUITMENT AND SELECTION

In this chapter, we will shed some light on the confidences of hiring and employing methods, used by the recruiters. Recruitment is approximately classified into two different categories − Internal Sources and External Sources.

➢ INTERNAL SOURCES OF RECRUITMENT:

Internal bases of employment refer to appointment employees within the group internally. In other words, candidates looking for the diverse positions are those who are presently employed with the similar organization.

At the time recruitment of teams, the initial contemplation should be given to those staffs who are presently working within the organization. This is an significant source of staffing, which delivers the chances for the growth and utilization of the current resources within the organization.

Internal sources of recruitment are the best and the coolest way of choosing resources as presentation of their work is previously known to the organization. Let us now debate more on the numerous internal sources of recruitment.

- **PROMOTIONS:**

Promotion refers to advancement the team of the employees by estimating their presentation in the organization. It is the procedure of shifting an employee from a lower position to a higher position with more tasks, remuneration, facilities, and status. Many organizations fill the higher vacant positions with the process of promotions, internally.

- **TRANSFERS:**

Transfer refers to the procedure of switching from one job to another without any alteration in the rank and duties. It can also be the shifting of employees from one department to another department or one location to another location, depending upon the requirement of the position.

Let's take an instance to know how it works. Take on there is a financial corporation called ABC Ltd. Having two branches, Branch-A and Branch-B, and an employee from Branch-A accepting from his job accountabilities. Therefore, this position has to be occupied for the continuance of the project with Branch-A. In this situation, as a substitute of searching or sourcing new candidates, which is time consuming and costly, there is an opportunity of shifting an employee from Branch-B to Branch-A, contingent upon the project necessities and the proficiencies of that respective employee. This internal shifting of an employee from one branch to alternative branch is called as Transfer.

- **EXTERNAL SOURCES OF RECRUITMENT:**

External bases of recruitment refer to hiring staffs outside the organization outwardly. In other words, the candidates seeking job chances in this case are those who are external to the organization.

External employees carry innovativeness and fresh opinions to the organization. Though hiring over external sources is a bit costly and tough, it has marvelous potential of driving the organization forward in achieving its goals. Let us now debate in detail the numerous external sources of recruitment.

- **DIRECT RECRUITMENT**

Direct recruitment mentions to the external source of employment where the recruitment of experienced candidates is done by assigning a notice of opportunity on the notice board in the organization. This technique of sourcing is also called as factory gate recruitment, as the blue-collar and technical workers are hired over this process.

- **EMPLOYMENT EXCHANGES**

As per the law, for definite job vacancies, it is compulsory that the organization offers details to the employment exchange. Employment exchange is a government object, where the particulars of the job seekers are stored and given to the bosses for filling the vacant positions. This external recruitment is cooperative in hiring for inexpert, semi-skilled, and skilled workers.

VII. CONCLUSION

An effective recruitment and selection process reduces turnover, we also get much better results in our recruitment process if we advertise specific criteria that are relevant to the job. Failure to recruit and select for the long term can result in high turnover. The recruitment and selection process is the time we not only identify a candidate who has the experience and aptitude to do the job that we are looking to fill, but also to find someone who shares and endorses our company's core values. The candidate will need to fit in well within our company's culture. The selection and recruitment process should provide our company with an employee who adapts and works well with others in our business.

REFERENCES:

1) *The Best Team Wins: Build Your Business Through Predictive Hiring – Adam Robinson*
2) *Hiring for Attitude – Mark Murphy*
3) *Who – Geoff Smart, Randy Street*
4) *High Velocity Hiring: How to Hire Top Talent in an Instant – Scott Wintrip*
5) *Human Resources Management – Aswarthapa*

CONSULTANCY MANAGEMENT – CRITERIA OF MARKETING STANDARDS

1.**GLN NAGARJUN**
Senior Consultant – TIBCO India Pvt Ltd
Hyderabad, India.
arjungollapalli@gmail.com

2.**A.MYTHREYI**
Software Engineer –Tech Mahindra IT Company, Hyderabad.
myinfo.co.in@gmail.com

3.**PNS VAISHNAVI**
Consultant – Deloitte, Hyderabad, India
vaishnavi0514@gmail.com

4. **A.SAI MITHRA**
Software Engineer - Tata Consultancy Services, Hyderabad.
saimithraagraharam97@gmail.com

ABSTRACT: The observation that impacts on management consultants help businesses improve their performance and grow by solving problems and finding new and better ways of doing things. It's not just in the private sector either many firms work with public sector organizations to help develop their services and, where necessary, reduce costs and make savings. As a consultant works as n a nutshell, and provide expert opinions, analysis, and recommendations to organizations or individuals, based on their own expertise. They're essentially fixers, serving as objective troubleshooters, and providing strategies to prevent problems and improve performance. Knowledge being democratized and information becoming more and more accessible to anyone, the role of management consultants is rapidly changing.

KEY WORDS: Consultancy, research and survey, organization growth, SWOT Analysis

I.INTRODUCTION: A consultant usually an expert or professional in a specified filed and has wide knowledge of the subject matter. The first is to provide expert advice in the areas in which we specialize. The experience of our professional consultant's means that are ideally placed to provide our clients with the knowledge, information and advice they required consulting or consultancy firm is a business of one or more experts (consultants) that provides professional feedback to an individual or an organization for a fee. The deliverable of a consultant is usually advice or a recipe to follow to achieve a company objective, leading to a company project. Consultant will be to work in partnership with clients, advising

them how to use information technology in order to meet their business objectives or overcome problems. It will work to improve the structure and efficiency of information technology systems in various organizations. Principle or practice, which is generally accepted between practitioners and academics as an integral part of the marketing concept and/or the marketing management processes, and conform to integrity and ethical conduct in the profession of marketing. *ref [1]*

II. **STANDARDS – TREND:**

Criteria regarding to estimating the Standards in Present Market

i. Information regarding the market and competition – estimating / predicting the situation,

ii. Analyzing the price criteria as a part of SWOT Analysis,

iii. Product / Service Management,

iv. Planning – Short term/ long term

v. Budgeting,

vi. Promotion - project to the client / customer,

vii. Enchasing the product for next attempt,

viii. Accepting the change or feedback for betterment

The standards will provide a framework and reference for training programs and marketing designations *ref [2]& ref [8]*

III. CONSULTATION & TYPES:

Expert Model: It is a model of telling, assuming and selling as per the client perception or giving expert service or information, the buyer is an individual manager or representative of some group in an organization it defines a needed and concludes that the organization has neither the resource nor the time to fulfill that needed. This model is almost by definition totally content oriented.

Assumptions:

- Whether or not the manager has correctly diagnosed his own needs
- Regarding the content communicated in proper way or not
- Requirements were accurately assessed the capabilities of consultant to provide the information service
- Consequences of having the consultant gather information or implementing the changes as per marketing standards
- Whether or not external reality than objective studied

Doctor - Patient Model: The core of this model is that client experience some symptoms that something went wrong but does not have due as how to go about figuring out what is wrong or how to fix it. This process is delegate completely to the consultant along with obligation to come up with a remedy, in this scenario client is totally depend up on consultant until such a time as the consultant make a prescription, unless consultant engages the client in becoming more active.

Assumptions:

- It is testing and supporting process after the deployment for resolving the issues
- Client has correctly interpreted the symptoms of sick ' area"
- Client has thought the consequences i.e., is willing to accept and implement whatever remedy given
- If the result is success the attempt is success

Process Consultation: It is the creation of relationship with the client that permits the client to perceive , understand and act on the process events that occur clients internal and external environment in order to improve the situation as defined by the client. *ref[3]*

Key Focus Areas of Process Consultation

- It build a relationship and permits consultant and client to deal with reality
- Remove the area of ignorance
- Acknowledge the consultant behavior's as being always on intervention
- Help the client to figure out what they should do about the situation

Assumptions

- Client not only needs help in making an initial diagnosis but it would be benefit from the participation in the process of decision making
- Constructive intent and some problem solving – logical mindset for approaching the problem solving issues
- That the client is ultimately the only one who knows what form of solution will work *ref [4]*

IV. **PRINCIPLES OF CONSULTATION**:

- Always try to gather the information past present for designing the future instance as per market standards
- Staying with reality avoiding the gossips and wastage
- Access to the ignorance and accepting the feedback
- Importance to the Time Management
- Problem and solution
- Continuity and should go with the flow
- Sharing the problem regarding the doubt and getting the " Clarity of thought" *ref [7]*

V. **ROLES AND RESPONSIBILITIES:**

- Conducting research, surveys and interviews to gain understanding of the business
- Analyzing statistics
- Detecting issues and investigating ways to resolve them

- Assessing the pros and cons of possible strategies
- Compiling and presenting information orally, visually and in writing
- Making recommendations for improvement, using computer models to test them and presenting findings to client
- Implementing agreed solutions
- Developing and implementing new procedures or training. *ref[5]*

VI. KEY SKILLS:

- Commercial awareness
- Good numerical skills
- Attention to detail
- Analytical skills
- Excellent interpersonal skills
- Tact and persuasive ability
- Team working skills
- IT skills
- Good oral and written communication skills
- Self-motivation *ref [6]*

VII. CONCLUSION: Consultants offer advice and expertise to organizations to help them improve their business performance in terms of operations, profitability, management, structure and strategy by basing marketing standards. Although the workload can be heavy, consulting is a sociable profession with plenty of networking opportunities. The work stretches across a variety of areas, including management, strategy, IT, finance, marketing, HR and supply chain management.

REFERENCE:

(1) Management study guide

(2) Strategic Management : A competitive Advantage Approach : Fred R David

(3) Valuation: Measuring and Managing the value of Companies

(4) Marketing Management – Philip Kotler

(5) Case interview secrets : A former McKinsey job offering for consultants

(6) A Management consultancy: Philip Sadler

(7) The Basic principles of Effective consulting : Linda K Stroh & Homer H. Johnson

(8) Consulting Success: The Proven Guide to Start the Consulting business by Michael Zipursky

CORPORATE SOCIAL RESPONSIBILITY AND EMPLOYEE ENGAGEMENT

RAJPUT SAI SUNIL SINGH
Student, Department of Business Administrations
MRCE – MBA,
Hyderabad.
Mail Id: r.sunilsingh97@gmail.com

Abstract

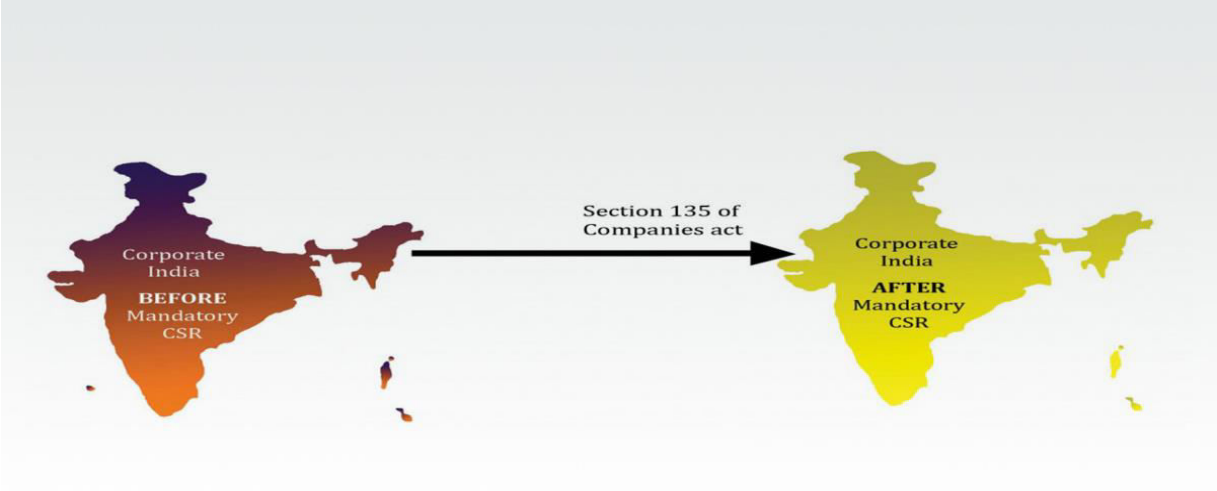
Corporate Social Responsibility is the way through which corporate companies address the every stakeholder. Corporate Social Responsibility moving far ahead from its age old domain of charity has now reached to a new mark of corporate responsiveness and action to social issues and sustainability in order to advance further towards a new era of collective future action for factoring the sustainable business strategy and development of the society. The purpose of this paper is to know the reasons for growth of CSR in India and to know what potential benefits of CSR to companies are. This paper also looks at a new framework for employee engagement and its practical applications and benefits for communities and corporations. Employee Engagement through service learning in executive development programs and innovative corporate-community volunteerism highlight how this can depend employees' identities as corporate citizens.

Introduction

In moderns days it not enough for businesses to simply buy and sell their products and services without considering the world in which they operate. The term "corporate social responsibility" came into existence in the late 1960s and early 1970s after many multinational corporations formed the term stakeholders, meaning those on whom organization activities have an impact. . It was used to describe corporate owners beyond shareholders as a result of an influential book by R. Edward Freeman, Strategic Management: Companies must voluntarily do business in an economically, socially and environmentally responsible manner to be sustainable for long term. Corporate social responsibility refers to business practices involving initiatives that

benefit society and its people. A business's CSR can encompass a wide variety of tactics, from giving away a portion of a company's proceeds to charity which benefits society to implementing "greener" business operations. The concept of CSR rests on the ideology of give and take. Companies take resources in the form of raw materials, human resources etc from the society. By performing the task of CSR activities, the companies are giving something back to the society (Companies Act 2013 (Companies Act) according to this act it is mandatory to every company to practice CSR in INDIA. CSR policy functions as a built-in, self –regulating whereby business monitors and ensures its active compliance with the power of the law, standards, and norms.

CSR in INDIA



India is the first country in the world to make corporate social responsibility (CSR) mandatory, following an amendment to the *Companies Act, 2013* in April 2014. Businesses can invest their profits in areas such as education, poverty, gender equality, and hunger as part of any CSR compliance.

The amendment notified in *the Companies Act, 2013* requires companies with a net worth of INR 500 crore (US $70 million) or more, or an annual turnover of INR 1000 core (US $140 million) or more, or net profit of INR 5 crore (US $699,125) or more, to spend 2 percent of their average net profits of three years on CSR.

Prior to that, the CSR clause was voluntary for companies, though it was mandatory to disclose their CSR spending to shareholders. CSR includes but is not limited to the following:

- Projects related to activities specified in the Companies Act; or
- Projects related to activities taken by the company board as recommended by the CSR Committee, provided

those activities cover items listed in the Companies Act.

The methodology of CSR

CSR is the procedure for assessing an organization's impact on society and evaluating their responsibilities. It begins with an assessment of the following aspects of each business:

- Customers;
- Suppliers;
- Environment;
- Communities; and,
- Employees.

The most effective CSR plans ensure that while organizations comply with legislation, their investments also respect the growth and development of marginalized communities and the environment. CSR should also be sustainable – involving activities that an organization can uphold without negatively affecting their business goals.

Organizations in India have been quite sensible in taking up CSR initiatives and integrating them into their business processes.

It has become progressively projected in the Indian corporate setting because organizations have recognized that besides growing their businesses, it is also important to shape responsible and supportable relationships with the community at large.

Companies now have specific departments and teams that develop specific policies, strategies, and goals for their CSR programs and set separate budgets to support them.

Most of the time, these programs are based on well-defined social beliefs or are carefully aligned with the companies' business domain.

Employee Engagement and CSR

Most companies today are doing something to engage their employees. The "war for talent" has seen employers pitch starting bonuses, flexible hours, challenging projects, the latest mobile technology, and amenities ranging from a concierge service

to health club memberships in hopes of attracting young talent and retaining "high-potentials" who might otherwise look elsewhere. This is certainly a step up from treating employees like expendable parts that could be replaced without a second thought. And it works: but only up to the point that another employer offers the same or better incentives or until the economy goes into recession and employers cut out the frills. Over the past few years, corporate social responsibility (CSR) has been heralded as a new "tool" to recruit and retain employees. Because (2006) Cone found that three of four of the generation (born 1978 to 1998) want to work for a company that "cares about how it impacts and contributes to society and its people." It is also found that, among those already in the workforce, nearly seven in ten say that they are aware of their employer's commitment to social/environmental causes, and 65 percent say that their employer's social/environmental activities make them feel loyal and responseble to their company.

Why Add CSR to Employee Engagement?

Why would a company concern itself with and seek to activate the people's identities as citizens of a corporation, community, society and group ? One reason is that when employees find that their company welcomes the full range of their interests and aspirations, including, for instance, a personal desire to serve society and/or protect the envoriment, they feel welcome to bring their "whole selves" into the workplace. This brings more commitment to one's work, a deeper connection to a company, and a broader sense of meaning associated with one's job and employment. As an example, leading CSR companies regard employees as important stakeholders who express their voices not only in employment and practices in nature, also on social issues related to employment. These firms have diversity councils, work/family forums, and associations of minorities, women, and gay/ lesbians, whereby employees can share their interests. These affinity groups not only provide input to and feedback on company policies and practices, but they also influence public positions taken by their firms.

- Witness, as an example, the increasing number of companies in the USA taking an stand on gay rights or joining forces with other firms in a coalition of businesses, Voices for Working Families, to promote work and balance family.
- A second reason is that when employees feel free to bring these multiple identities into the working place , they become a microcosm of the markets and societies in which a firm operates. Recognizing the, top companies like IBM in its Innovation and Nokia in its World Map exercise, regularly consult with employees on social and factor their ideas not only employment policies, but also into corporate social responsibility, business innovations, and their overall socio-commercial agenda.
- A third reason is that employees their identities, whether enhanced or diminished by their companies, into society and the market. Studies find that the prime source of information about the citizenship of companies comes via word of mouth. Employees wants to live and work responsibly and fulfill through their companies thereby serve as effective

brand ambassadors for their firms through their word-of-mouth commentary. They also produce social capital—a web of positive relevant relationships—that connects their companies to other stakeholders and the public at large.

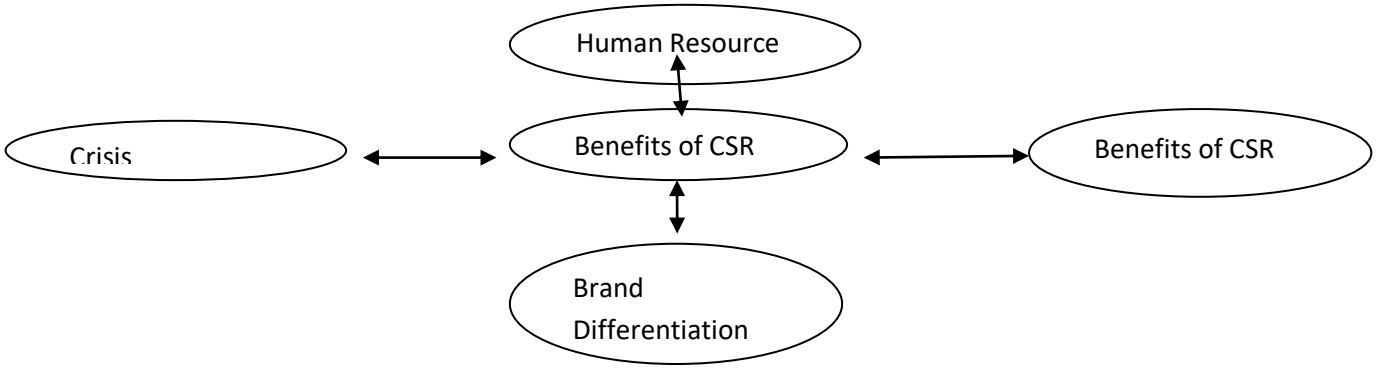
Finally, employees who feel empowered as citizens produce social value through their volunteer service, their jobs, products, and services, and the enriched understandings of corporate citizenship that are shared with friends, a subject of debate among colleagues and critics, and ultimately passed on to their children.

Engaging Employees Through CSR: Volunteerism

Leading companies in world actively engage every employees in CSR programs and often this takes the form of volunteerism, especially relevant for Indian employees, but of increasing interest to working people around the world.

- Cause effective configuration:
  Focus on causes to the business; leverage skills and other company assets (commercial products); partner where appropriate with a nonprofit service arm. This makes the community service seem "serious" rather than a "day off " to the employees.
- Strategic business positioning:
  Integrate volunteerism into the larger CSR thrust of the company; connect to business goals; manage the volunteer service effort professionally. This ensures that employee volunteerism "makes a difference."
- Sufficient investment:
  Add treasure to talent.
- Culture of engagement:
- Encourage and support volunteerism; senior management of companies modeling; middle management buy-in. This makes community service something that "we all do."
- Strong participation:
  Large numbers; paid time off; media coverage. This sends a message that volunteerism matters.
- Actionable evaluation:
  Track numbers, hours; feedback from volunteers and those they serve; business and social outcome measurement. Excellence matters, too, not just "showing up."

Potential business benefits of CSR



The nature of benefits of CSR for any organization can vary depending on the nature of the enterprise. The business case for CSR within a company will likely rest on one or more of these arguments.

1. Human resources
   A CSR programme can help to recruitment and retention of humane resource. Potential recruits often ask about a firm‘s CSR policy during an interview, and having a comprehensive policy can give an advantage.
2. Crisis management
   Managing crisis is a important part of many corporate companies strategies. Reputation take decades to build up and can be ruined in hours through incident such as corruption or environmental accident. These can draw attention from regulators, courts, governments & media. Building a genuine culture of doing the right thing within a corporation can offset these risks.
3. Brand differentiation
   In large market places, company‘s striving for a unique selling point. CSR affects brand differentiation through customer preference and employee engagement. CSR also helps in building customer loyalty based on ethical values.
4. License to operate
   Corporation are keen to avoid interference in business through taxation. By taking voluntary steps, they can tell governments that they are taking issues such as health & safety, diversity, or the environment seriously as good corporate citizen for positive impact on the environment. CSR remains essential to building and maintaining a strong brand image and goodwill. A CSR initiative also helps in protecting a company from harm during a controversial or challenging event

Conclusion

CSR have no boundaries and are not constrained by race, color, or religion. Sadly, concern for the community is often mistaken for socialism. On the contrary, every citizen is an asset in economic activity and has opportunities to succeed. This invisible culture can shape brighter future for nations. Organizations must realize that government alone will not be able to get success in its endeavor to uplift the down trodden of society. The present societal marketing concept of companies is constantly evolving and has given rise to a new concept- CSR. Many of the leading corporations across the world had realized the importance of being associated with socially relevant causes as a means of protecting the goodwill and reputation, defending attacks and increasing business competitiveness. It stems from the desire to do well and get self-satisfaction in return as well as societal obligation of business. There is also budding interest in companies (and certainly among employees) in what might best be called "socially responsible jobs." Long ago employers learned of the advantages of job enrichment—whether in the form of more variety, autonomy, and challenge or in opportunities for influence over how the job is done. Employee engagement via CSR enables employees to affirmatively answer some of these questions. A study by the National Environment Education Foundation (2010) finds that employee education and engagement efforts aimed at "greening" their companies have paid big dividends.

References

Article 13. Patagonia. (2006, November). www.article13.com/A13ContentList.asp?strAction =GetPublication&PNID=1245

Boccalandro, B. (2009). Mapping success in employee volunteering. Boston: Boston College Center for Corporate Citizenship

Cone. (2006, October 24). The 2006 Cone millennial cause study. Boston, Author. www.coneinc.com. Gallup Employee Engagement Index. www.gallup.com/consulting/52/employeeengagement.aspx

Macey, W.H., & Schneider, B. (2008). The meaning of employee engagement. Industrial and Organizational Psychology, 1, 3–30

.

• Jamali, D., and Mirshak,R.(2010).Business-conflict linkages: Revisiting MNCs, CSR, and conflict, Journal of Business Ethics, 93(3), pg 443

Clark, William H., and Babson, Elizabeth K. (2012). Business Organizations: When "Business Purpose" Disappears; How Benefit Corporations are Redefining the Purpose of Business organizations. William Mitchell Law Review, Vol. 38, pg. 817-844

OCCUPATIONAL STRESS AMONG EMPLOYEES EMPLOYED IN VARIOUS INDUSTRIES- A LITERATURE REVIEW APPROACH

Dr.N.MUTHUKUMAR,
Professor& Dean ,
Department of Business Administration,
Malla Reddy College of engineering,
Hyderabad.
muthumba_2005@yahoo.co.in

Mrs.Dhanalakshmi komirisetty
Assistant professor ,
Department of Business Administration,
Malla Reddy engineering College ,
Hyderabad ,
India,
ms.dhanu1144@gmail.com

Dr.C.Kathiravan
Associate Professor,
Department of Business Administration
Annamalai university
Tamilnadu

Dr.S. YAVANARANI
Associate Professor,
CMS Business School, Jain university
Bengaluru
India
s.yavanarani@gmail.com

ABSTRACT

Stress is unavoidable in the present situation. It isn't limited to one's who feel it rather has become an open issue and is looking for consideration from individuals around. Every individual hold diverse view about worry as it very well may be experienced from assortment of sources. Indian advanced education segment is moving towards instruction industry where private players are thinking of innovation based instructive procedures. Instructing is naturally a distressing occupation, and by numerous records, it's getting all the more so. Research has shown that police pressure antagonistically influences the official's social emotionally supportive networks, for example, marriage and family and other social connections.

Key words: occupational stress, stress, burnout, emotional intelligence, work life balance

1. Introduction

Stress is unavoidable in the present situation. It isn't limited to one's who feel it rather has become an open issue and is looking for consideration from individuals around. Every individual hold diverse view about worry as it very well may be experienced from assortment of sources. However, worry for pressure has expanded these days with expanding worry about work life parity and personal satisfaction. Indeed, even the quick changes are found in the social arrangement of Indian culture where progressively joint families are separating into family units and there is a weight of family obligation on one shoulder. Urbanization, station framework, political impacts, and quick industrialization are for the most part assuming its job in expanding distressing conditions for the network on the loose. Cases for pressure are twice as high as those paid for non stress physical damage at the work place, acquiring a yearly cost of about $200 billion as cited by Agrawal (2001). A gigantic weight is made upon the people because of blend of way of life issues and in general ecological pressure prompting ceaseless diseases like coronary episodes, expanding blood pressures and mental issues like sadness and suicides. These issues are intently interlinked to pressure. It is very much recognized that downturn impacts the event of different infections. In India, around 57% of the individuals are influenced by wretchedness according to the distribution by WHO. The psychological weight that is made because of different powers impacting an individual is offering ascend to burdensome issue, uneasiness issue, sedate use issue, schizophrenia. PM Narendra Modi in his discourse through Mann ki Baat Radio on March 27, 2017, encouraged Indians to discuss melancholy and look for help if necessary. It isn't amazing that pressure has raised dynamically in the course of recent decades stamping it as a subject hard to keep away from. One needs to give due consideration regarding pressure and its negative results.

Because of changes in the idea of work and expanding rivalry, individuals need to manage expanding business related pressure which is felt because of the distinction between the requests and weights at the work place and the human propensity to adapt to it. Great associations are accepting this issue as a test and are contemplating the repercussion it might bring. Worry in the working environment is basic all through the world in each industry and every one of the associations. There are number of stressors that influence the presentation and character of a person. Change in the workplace may fuel worry for certain people while some might be influenced by the advantages. Down to earth information and accessible writing for various divisions like Banking, IT, Healthcare and Education pointed worry as one of the most persuasive factor influencing the fulfillment level of the representatives at the working environment

2. Occupational Stress among university teachers and school teachers

In the ongoing period of expanding rivalry, Indian advanced education segment is moving towards instruction industry where private players are thinking of innovation based instructive procedures. Instructing is naturally a distressing

occupation, and by numerous records, it's getting all the more so. Understudies carry the dangers of innovation to the homeroom and its getting troublesome nowadays to keep power over the equivalent. Study hall conduct is reliably refered to as one of the principle drivers of instructor worry alongside the remaining task at hand. Scientists have reliably inferred that showing is a distressing occupation, and that a critical number of educators, maybe even a dominant part, are influenced by business related worry as closed by Rudow (1999); Dunham (1998); Kyriacou (1989). Instructor stress might be characterized as a reaction disorder of negative impact coming about because of parts of the educator's activity and interceded by the observation that the requests made upon the instructor comprise a risk to their confidence or prosperity as proposed by Kyriacou and Sutcliffe (1978) in their investigation.

Indeed, even work life balance among educators has gotten perhaps the best challenge in the present time. The associations managing in higher learning are quickly incorporating the trend setting innovation with the assistance of ICT into showing learning process. The unexpected change in the requests to keep up the instructive limits through innovative headways, is putting overwhelming weight on training experts making awkwardness in their family life. Because of such requests, the idea of Work life balance has gotten vital for instructing experts to adapt to the unsure condition in the instructive set up. Instructors are required to spend additional hours to be powerful in their job. This calling doesn't just concentrate on instructing through ones specialized information yet in addition requests a solid hold over delicate aptitudes and fundamental abilities which can cook towards the satisfaction of instructive objectives. On the off chance that educators are focused and incapable to adjust their work and family, their presentation in class will be influenced and eventually it'll impact the mass. As per Guest (2002), the explanations behind the ascent of concern with respect to work-life balance are the weight and heightening of work, expanding center around nature of home and network life and the frames of mind and estimations of individuals. This thusly prompts huge ascent in medical issue, which thus monetarily influence both the business just as the administration as recommended by Frone et al. (1997).

In numerous nations educating is frequently considered as one of the most unpleasant calling. As per Kyriacou (2000) educators stress can be characterized as "the experience by an instructor of unsavory negative feelings, for example, outrage, dissatisfaction, uneasiness, misery and apprehension, coming about because of some part of their work". Kyriacou and Schutcliffe characterized educators worry as, " a reaction disorder of negative impacts, (for example, outrage or sadness) typically joined by possibly pathogenic physiological changes, (for example, expanded pulse) coming about because of parts of the showing work and intervened by the recognition that requests made upon instructor comprise a risk to his/her confidence or prosperity and by methods for dealing with stress

actuated to diminish the apparent threat." Vandenberghe (1999) states that, instructors stress is the "general term to portray negative feelings of educators that are reflected in aversive requests to their work".

Banerjee, S., and Mehta, P. (2016) in their investigation worked after securing the forerunners of position pressure and their impact on work execution as far as both disappointment and shirking among the employees. The information was gathered from 110 staff educating in B-Schools in Andhra Pradesh. Factor Analysis was performed to distinguish significant factors influencing feeling of anxiety and Regression examination was done to discover the effect of weight on work execution. The outcomes uncovered that instructing pressure prompts work evasion, while work over-burden pressure and poor relational connections lead to work disappointment.

Luhar U., Dr. Vaghela K. (2016) in their investigation attempted to quantify the feelings of anxiety among the instructors and the police officers. Since educators stress has been a worry in the ongoing time and the police officers likewise need to work for the prosperity of the residents nonstop, these two divisions were chosen with the end goal of study. 30 instructors and the 30 cops from Amreli locale were grilled with the assistance of a poll confined by utilizing Occupational Stress scale by A.K. Shrivastav and Singh (1981) and Job Satisfaction scale by Parikh (1985). T-test was utilized as a factual measure to arrive at the resolution. The outcomes demonstrated that there is a noteworthy contrast between the degrees of worry as experienced by the educators and the individuals from the Police office. Additionally it was discovered that there is a distinction in the activity fulfillment level among both the offices.

Sabherwal, N., Ahuja, D., George, M., and Handa, A (2015) directed an investigation on word related worry among employees in Higher Education Institutions in Pune. Their investigation comprised of (N=200) educators of various higher instructive foundations in Pune. The aftereffects of the investigation demonstrate that absence of standard breaks (85%) and long working hours (83%), provocation by chiefs/staff/understudies (75%), absence of correspondence with staff (73%), poor compensation possibilities (81%) pace and force of progress (75%), high level of vulnerability about work cause greatest pressure. Additionally Age, Gender and Marital status impacts the degree of stress looked by the employees and the theory of work requests, connections at work, work jobs, work changes and bolster related components influence feeling of anxiety among personnel was acknowledged.

Visotskaya, N., Cherkashina, E., Katcin, O., and Lisina, L. (2015) in the examination entitled Studies on University Professors' Emotional Burnout' inspected passionate burnout disorder among the University educators of the Department of Transport in Siberian Federal University. Distinctive indicative tests and surveys were utilized to uncover the level of passionate burnout and expert burnout among 22 instructors managing

philanthropic and designing subjects. Mann Whitney U-test was applied to discover the contrast between the gatherings under examination in connection to the qualities of their expert action. Side effects of pressure improvement were ordered into 3 phases of Tension, Resistance and Exhaustion. The outcome uncovered that the passionate burnout manifestations among the speakers of compassionate subjects at the pressure' organize are distressing experience, feeling cornered, nervousness and misery where as at the obstruction' arrange incorporates deficient enthusiastic response and passionate and moral bewilderment. Passionate shortfall, enthusiastic and individual separation is common at the weariness' arrange prompting encountering more passionate burnout than instructors of building subjects. It was additionally discovered that each phase of passionate burnout is either in procedure of development or it is as of now shaped and is because of expert exercises of the subjects.

An examination on the components affecting occupation fulfillment of MBA employees in Gujarat state‖ was directed by Dave, N., and Raval, D. (2014) with the goal of estimating the legitimacy of different elements that influence the activity fulfillment of the school and college educators of Gujarat. It additionally planned for recognizing the components that influence the educators at individual, gathering and authoritative levels. The example for the examination comprised of 82 employees having a place with 25 MBA Colleges and University branches of Gujarat acquired through straightforward irregular testing. Different writings were alluded so as to comprehend different components influencing the activity fulfillment. The variables distinguished were isolated into Individual and Institutional components class. The data was gathered from the Faculty individuals with the assistance of Questionnaire and Interview strategy. The outcome distinguished the Job Satisfaction Factors (JSF) that influence the fulfillment level of the Faculty Members. The individual components incorporate fulfillment from study hall educating, Training and Faculty Development Programs, Performance Appraisal, Cooperation and conduct of friends (partners). Aside from this the institutional variables are physical working condition, understudy collaboration, understudies IQ, understudy interest, enthusiasm to learn, acknowledgment for additional work/subjective work, goals and unmistakably characterized advancement approaches, compensation, cooperation in basic leadership, the board style/the board reasoning/vision/strategic/at top administration, testing and intriguing work, employer stability, association culture, support for inquire about condition in association and notoriety of association in showcase. It tends to be henceforth inferred that the association and the top administration ought to guarantee that the approaches and practices in the association deal with these components to guarantee the staff work fulfillment which could eventually influence their degree of execution and educating.

Parveen (2013) directed an exploration on ―Faculty Stress in a Saudi Government University‖ with the goal of investigating the staff recognition towards word related

worry in the school of business. Information was gathered from 160 employees of different offices utilizing Faculty Stress Index (FSI). Multivariate investigation of Variance was led to distinguish and look at singular pressure contrasts as far as sexual orientation, scholarly position, nationality, scholastic degree, office, business status and No. of years working in college. The outcome indicated that the male individuals with a mean score of 2.3750 are having an excessive amount of weight than female personnel with a mean score of 1.8657 identified with Students' Interaction Subscale. It was additionally noticed that age and conjugal status doesn't have any impact on male and female employees. When all is said in done, huge contrasts in impression of individual worry as far as statistic factors were watched

In a review of rustic and urban auxiliary instructors, Abel and Sewell (1999) found that urban teachers experienced fundamentally more worry than provincial teachers in with respect to poor working conditions and staff support. In the two kinds of schools, understudy trouble making and remaining task at hand were found to most critical indicators of worry in the two sorts of schools. As indicated by Kim-wan (1991) less experienced instructors and those with less social help have higher burnout.52 Travers and Cooper (1997) study discovered remaining task at hand and pay as essentially connected with word related worry among English educators.

In an example of auxiliary teachers in Netherlands, Brouwers et al (2006) found that absence of social help among instructors was influencing educators self-adequacy conviction and causing burnout. Then again burnout was found to anticipate degree of saw absence of social support.54 Male and May (1997) found inordinate outstanding task at hand (45%) as the most unpleasant in educators of kids with a custom curriculum needs. This was trailed by administrative work (41%) and testing conduct (21%)

An investigation of 1000 understudy instructors Morton et al, (1997) and Lewis (1999) revealed understudies conduct altogether connected with worry among educators. Lewis (1999) additionally detailed that concerned instructors were falling wiped out because of stress.50 Similarly, Axup (2008) et al found that understudies conduct was huge reason for tension among educators.

Guthrie (2006) found that female educators in Australia were having higher business related worry than their male partners.

3.Occupational Stress among police personnel

A larger part of research has shown that police pressure is basecL on the individual cop's view of a specific occasion. Truth be told, Violanti (19831) proposed that officials change their discernment as their administration time increments. Sadly, the upsetting occasions experienced by singular cops influence them behaviourally, mentally, socially, inwardly, and physically.

Research has shown that police

pressure antagonistically influences the official's social emotionally supportive networks, for example, marriage and family and other social connections. A broad writing survey by Alkus and Padesky (1983) concerning pressure and cops showed that conjugal challenges are regularly announced and that separation rates, especially during the initial hardly any long periods of police administration, show up exceptionally high. In addition, they showed that conjugal issues seem, by all accounts, to be the most critical hastening worry in cop suicides, particularly for more youthful officials.

Police pressure is likewise influenced by the official's age, conjugal status, and instruction level. Research led by Spielberger et al. (1981) showed that more youthful officials (ages 18-29) evaluated "Court Leniency," "Insufficient Salary," and "Family Demands" as more exceptionally distressing than did more seasoned officials. More seasoned officials appraised "Absence of Recognition" as exceptionally unpleasant. Mayes, Barton, and Ganster (1991) looked into the directing impacts of age on stressor-strain connections in an example of 523 cops, firemen, circuit testers, and directors, ages 18-63 years. They found that age directed the connection between different occupation stressors and physiological (i.e., adrenaline and wellbeing side effects) and mental strains (i.e., misery and life fulfillment).

Spielberger et al. (1981) likewise found that officials with the most elevated level of training (post-graduate work) appraised "On-the-Spot Decision Making," "High Moral Standards," and "Over the top Paperwork" as significantly less distressing, and "Weakening Injury" as substantially more unpleasant than officials with either a secondary school, some school, or higher education level of instruction. Actually, Burke (1994) found that cops with more instruction revealed less psychosomatic indications and negative inclination states. As to status, Spielberger et al. (1981) found that solitary officials appraised "Physical Attack" as increasingly unpleasant and "Advancement" as considerably less distressing, while isolated or separated from officials evaluated "Absence of Recognition" as profoundly upsetting. Hitched officials evaluated "Fast Chases" as increasingly unpleasant and "Physical Attack" as less distressing.

Ethnicity and sexual orientation somewhat influence worry in policing. Wexler and Logan's (1983) examination of worry among female cops showed that negative demeanors of male officials, preparing, introduction to catastrophe and inconvenience, bunch fault, and bits of gossip were their significant wellsprings of stress and that the activities of male cops expanded the anxiety for ladies officials

Janik and Kravitz (1994) looked into the records of 134 cops who had as of late experienced a qualification for-obligation assessment dependent on questions with respect to their capacity to suitably release their obligations under distressing conditions. They found that 55% of the officials admitted to past suicide endeavors. Their examination further

demonstrated that officials revealing conjugal issues were 4.8 occasions bound to have endeavored suicide and 6.7 occasions more probable in the event that they had been suspended.

Anderson (1995) contemplated pressure and its connects among 216 cops. He found that help from the official's companion, division, and social condition was unequivocally connected with physical wellbeing, incessant nervousness, mental alteration, and the impression of seriousness in word related stressors. Officials who were happy with the help they got were bound to report better physical wellbeing, less post-horrendous pressure issue symptomatology, lower levels of nervousness and enthusiastic depletion, and less worry due to hands on stressors than did officials who were not happy with their emotionally supportive networks.

Myendeki (2011) demonstrated that when the word related stressors, absence of assets happen cops actualize shirking adapting (COPE) as a system, which prompts the experience of Exhaustion. Results additionally show that cops utilize dynamic adapting, intellectual adapting and going to religion as methods for directing the pressure burnout relationship. Results additionally show that the adapting procedure shirking adapting, utilized by male officials lead to criticism. At the point when female cops experience work requests and an absence of assets they utilize looking for enthusiastic help as an adapting system.

Glass et al. (1993) suggested that the issue of apparent employment control assumed a critical job on the effect of enthusiastic fatigue on sorrow, with depleted medical attendants seeing themselves as less engaged. Psychosocial stress results in biophysical reactions like raised pulse, expanded circulatory strain, expanded muscle pressure, expanded corrosive discharge (Anderson et al., 2002) and mental concerns like burnout and weakness (Harpold and Feenster, 2002). Archibauld et al. (1999) found huge relationships between's degrees of passionate depletion, individual achievement, and indications of stress-related sickness just as its harming consequences for the physical and enthusiastic working. Educator burnout additionally influences instructing (Wisniewski and Gargiulo, 1997), responsibility to showing calling Farber (1984) and prompts negative elucidation of understudy conduct (Whiteman et al., 1996)

4.Occupational Stress among and banking employees

Enekwe, C. I., Agu, C. I., and Nnagbogu, E. K. (2014) planned for distinguishing the components causing word related worry among the bank workers and stress the executives systems utilized by male and female representatives of Nigeria banking industry. ANOVA was utilized as an exploration procedure. The specialists from the t-test investigation found that male and female bank representatives experience comparative sort of stress and stress the executives isn't sexual orientation touchy so the pressure the board strategies among male and female workers don't vary fundamentally. Analyst recommended different methods like preparing, all out computerization and others to empower bank representatives' adapt to

the degree of weight on the normal premise and not simply to offer during the difficult circumstances. They further proposed conduct procedures, unwinding systems and subjective strategies to oversee pressure.

The investigation entitled Occupational worry in banking division' by Dhankar, S. (2015) was attempted to decide the degree of stress experienced by the individuals and furthermore to examine the effect of different segments of worry among the representatives of 20 banks of Kurukshetra, Panipat, Sonipat and Karnal district. The investigation was directed among 200 workers of both open and private division banks. The study was finished with the assistance of word related pressure estimation scale comprising of 46 things. The consequence of the examination demonstrates that the private area workers feel worry because of the Role over-burden though the open division representatives feel more worry because of Unreasonable gathering and political weight. No distinction was found between the various constituents of word related worry in private and open banks. The general score of word related pressure found in private banks is 49.71% and 50.28% out in the open banks demonstrating that there is no distinction in the degree of stress experienced by the workers of the two parts.

Samartha, V., and Begum, M. (2014) in their investigation entitled ―A similar examination of word related worry among the workers out in the open and private part banks in Dakshina kannad district‖ planned for finding the effect of word related weight on the representatives of private and open area banks. A review was directed among 537 workers and different estimates like Chi-square test, relapse examination and factor investigation were utilized to gauge the outcomes. The outcome demonstrated absence of productive labor and execution pressure as the most upsetting components pursued by work conditions, requests of the family, undue desires from work, unexpected possibilities and occupation inflexibility among the open segment bank representatives while in private segment banks, nonappearance of worker contribution, physical condition and severe association exercises, versatility to change are among the elements affecting the feeling of anxiety of the representatives. It was likewise conjectured that representatives of both open division and private area banks experience the comparative degree of stress.

Pradhan, R., and Tomar, P. (2013) in their examination entitled ―Evaluating worry in the Indian Banking scenario‖ researched the degrees of stress experienced by 60 representatives of 3 diverse private division and open area banks. The examination was planned for understanding the variables that are causing unpleasant circumstances among the workers. A correlation was likewise done to distinguish which of the two parts offer better pressure the executives techniques. The outcome after use of Chi-square testing showed that the pressure is overseen successfully in private area banks (ICICI) contrasted with other open division banks like SBI and BOB. The significant sources adding to the worry

among open area banks were the absence of the executives support, absence of meeting and correspondence, an unsupportive director, dread towards the board, working condition, and others though central point in private part incorporates an excessive amount of weight, ridiculous cutoff times, associations with colleagues, social foul play, aggressive institutional culture and others. The specialists proposed different procedures to deal with worry after the examination.

Devi, An., and Sharma, J. (2013) explored the job worry among the forefront bank representatives in Jammu and Kashmir. The examination was led among 501 representatives chose through irregular testing strategy. A size of 22 things was set up for the survey and connection strategy was utilized to check the relationship of the factors. So as to check the sufficiency of information for Factor Analysis, the Kaiser Meyere Olkin (KMO) measure was applied and the resultant score of >0.60 demonstrated that information is satisfactory for factor examination. The outcome uncovered that the representatives can be classified under three sections of ―overloaded employees‖, ―unclear workers and ―underutilized representatives, in view of their experience of job stressors. The consequences of factor investigation demonstrated job vagary, job abundance, job obtrusiveness, job difference, job enlargement, self-lessening, job stronghold and asset lack as the main considerations adding to the degrees of worry among the representatives individually.

Shukla, H., and Garg, R. (2013) directed research with the point of distinguishing the reasons of worry among the bank workers and their endeavors to adapt to the pressure. The Data was gathered through self-organized poll from 50 representatives of different nationalized bank arranged in Indore with the assistance of Percentage Analysis technique. It was gotten from the outcome that lion's share of the representatives stays in worry because of different reasons including absence of value, work over-load, nonachievement of the objectives being most compelling ones. Representatives attempt systems like Yoga to ease pressure. There were not many workers who found the procedures utilized by bank to oversee pressure viably.

Ahmed, An., and Ramzan, M. (2013) directed an exploration so as to recognize the connection between work pressure and occupation execution on representatives of banking segment in Pakistan. A nearby finished poll was circulated among the 144 senior alumni workers falling under the classes of FTE (Fixed Tenure Employees) bank agreement and outsider agreement representatives. The factual instruments like relapse and connection was utilized to legitimize the targets. The Croan-bac Alpha was utilized to test the unwavering quality of the needy and autonomous factors which scored 0.694 and 0.637. The outcome indicated the negative connection between work pressure and employment execution. With the assistance of writing, the scientists recommended to expand the executives bolster and give hierarchical help to

diminish the degree of stress.

Adikaram, D. S. R., and Jayatilake, L. V. (2016) broke down the effect of work life balance on representative employment fulfillment in private division business banks of Sri Lanka. Information was gathered from 150 representatives of various business banks and was investigated with the assistance of connection and relapse. In the wake of applying the tests, no noteworthy relationship was found between working hours and working conditions and worker work fulfillment while a critical connection between work pressure, change of occupation and work-life balance projects and representative employment fulfillment.

Sarwar, An., and Aftab, H. (2011) in the title ―Work Stress and Family Imbalance in Service Sector of Pakistan‖ underlined the circumstances and end results of work pressure and Family irregularity issue. Specialists examined the reasons of worry among 500 center level chiefs working in banks alongside its effect on people and their relatives. The poll was framed and appropriated by multi organize irregular testing. Different tests like T-Test, Correlation, Man-Whitney U test, Kruskal Wallis test and Regression examination were applied so as to reach to the resolution. Individual and family effect of pressure was practically same in every one of the classes of age and no. of employments exchanged. Through connection investigation it was discovered that there is a solid positive relationship between's work pressure and work pressure family sway. Relapse examination indicated that 64.8% change in pressure sway on individual and irregularity in the family was a result of ―Work Stress.

5. Occupational Stress among nurses

The nursing calling is related with a few activity related requests that assume a job in making unpleasant workplaces. Studies have indicated that psychosocial stressors among attendants produce significant employment stress, which can prompt wellbeing issue and diminish the nature of nursing care (Kawano 2008; McVicar 2003).

Healy, McKay (1999) to distinguish what the medical caretakers saw as the significant reasons for worry in the work environment. Results indicated that the medical caretakers appraised their remaining burden as profoundly upsetting as far as both recurrence of its event and its apparent impact upon themselves. True to form, more elevated levels of revealed nursing pressure were related with lower levels of employment satisfaction.Stress tolerant attendant chiefs with high toughness levels detailed 35% less wiped out hours than their low solidness partners (Judkins, Masse and Huff, 2006). Stress tolerant medical attendant supervisors have exhibited less incessant utilization of evasion and cautious adapting systems and normally announced the impression of significant levels of family support (Judkins, 2001). The significance of social help in the working environment is additionally obvious in the writing and has been found to identify with expanded strengthening, expanded inspiration, and diminished employment strain (Shirey, 2004). To improve the individual and expert results of the job, nurture supervisors have explicitly recognized the

requirement for more help from senior organization in managing job changes and difficulties (Thorpe and Loo, 2003), more power and regard steady with expanding medical caretaker chief duties (Suominen et al., 2005)

6.CONCLUSION :

Word related Stress is an ongoing idea in root, yet a major test and a disturbing issue for representatives in every one of the areas. It is asserted that pressure is inescapable and if not controlled can influence the wellbeing and prosperity of the educators. By making mindfulness, taking restorative measures and vital pressure preventive exercises, the side effects and perils of pressure can be limited. Representatives who were skilled at utilizing and dealing with their feelings in the working environment were less inclined to report experiencing sick wellbeing (mental and physical) and were bound to report significant levels of employment fulfillment and hierarchical responsibility. These synergistic elements assume a significant job in encouraging the vital commitments nurture chiefs make toward building the sound workplaces that are vital for holding medical attendants in the workforce and for delivering positive basic leadership and wellbeing results for the two people and foundations. The connection between faculty assets and occupation fulfillment is more grounded among medical attendants that have an advancement centered objective. Medical caretakers with a counteractive action centered objective don't benefit from work force assets as much as attendants that have advancement centered objectives. The effect of choice expert on physical grievances is likewise more grounded among medical attendants with an avoidance center

REFERENCES

Ashoksinhji, J. H. (2018). Exploring the effects of Occupational Stress on Work Life Balance A Study of University Teachers of Gujarat.

Benner, P. E., & Wrubel, J. (1989). The primacy of caring: Stress and coping in health and illness. Addison-Wesley/Addison Wesley Longman.

Behdin, N. (2013). Quality of work life: investigation of occupational stressors among obstetric nurses in Northeastern Ontario (Doctoral dissertation, Laurentian University of Sudbury).

Bawa, N. (2012). A study of burnout in relation to occupational stress self efficacy hardiness and coping strategies among police officials.

Burnett, H. J. (2001). A study of the relationship between police stress and moral reasoning, coping mechanisms, and selected demographic variables.

Shirey, M. R. (2009). Stress and coping in nurse managers: A qualitative description (Doctoral dissertation).

Shyam Swaroop, S. (2008). Occupational stress among teachers working in secondary schools of karauli district of rajasthan state (Doctoral dissertation, SCTIMST).

Gardner, L. (2005). Emotional intelligence and occupational stress (Doctoral dissertation, Swinburne University of Technology).

Tennant, C. (2001). Work-related stress and depressive disorders. Journal of psychosomatic research, 51(5), 697-704.

ABOUT MRGI

Malla Reddy group of Institutions is one of the biggest conglomerates of hi-tech professional educational institutions in the state of Telangana, established in 2001 sprawling over 200 acres of land. The group is dedicated to impart quality professional education like pharmacy, Engineering & Technology, MCA, MBA courses. Our sole objective is to turn out high caliber professionals from those students who join us.

ABOUT MRCE

Malla Reddy group of Engineering (Formerly CM Engineering College) has been established under the aegis of the Malla Reddy Group of Institutions in the year 2005, a majestic empire, founded by chairman Sri Ch.Malla Reddy Garu. He has been in the field of education for the last 23 years with the intention of spearheading quality education among children from the school level itself. Malla Reddy College of Engineering has been laid upon a very strong foundation and has ever since been excelling in every aspect. The bricks of this able institute are certainly the adept management, the experienced faculty, the selfless non-teaching staff and of course the students.

ABOUT ICTIMES

ICTIMES started long back with its banner to promote the vision of future technologies that change the trends of life on this planet earth. Under this banner, the Department of Humanities Sciences and Management at MRCE organizes the ICAHSM – International Conference on Advances in Humanities Sciences and Management to provide a scholarly platform to ignite the spirit of Research and bring out the latent potential in teaching fraternity and student community. ICAHSM accommodates major areas like, English Language, Literature, Phonetics, Mathematics, Physics, Chemistry and Management Sciences.

ABOUT ICAHSM

International Conference on Advances in Humanities Sciences and Management (ICAHSM-2019) will bring together innovative academicians, researchers and industrial experts in the field of Humanities Sciences and Management to a common forum. The idea of the conference is, for the scientists, scholars, engineers and students from the Universities across the world and the industry as well, to present ongoing research activities, and hence of foster research relations between the Universities and the industry with the rapid development of trends and studies in the fields concerned. ICAHSM-2019 will provide a heartwarming platform to researchers, scholars, faculty and students to exchange their novel ideas face to face together.

Estd :2005

MALLA REDDY COLLEGE OF ENGINEERING

Approved by AICTE - New Delhi, Affiliated to JNTU - Hyderabad,Accredited by NBA & Accredited by NAAC.
ISO 9001:2015 Certified Institution, Recognition of College under Section 2(f) & 12 (B) of the UGC Act,1956.

Address: Maisammaguda, Dhulapally,(Post Via Kompally),Secunderabad - 500 100.
Ph: 040-64632248, 9348161222 ,9346162620. Email: principal@mrce.in
Website : www.mrce.in