IJCER

International Journal of Computational Engineering Research (IJCER)

Impact factor: **1.145**

IJCER Publishes Online and Print Version Both

Open Access

www.ijceronline.com

Editorial Board

Meisam Mahdavi
Qualification: Phd Electrical and Computer Engineering
Affiliation: University of Tehran, North Kargar st. (across the ninth lane), Tehran, Iran

Dr. Ahmed Nabih Zaki Rashed
Qualification: Ph. D Electronic Engineering
Affiliation: Menoufia University, Egypt

Dr. José M. Merigó Lindahl
Qualification: Phd Business Administration
Affiliation: Department of Business Administration, University of Barcelona, Spain

Dr. Mohamed Shokry Nayle
Qualification: Phd, Engineering
Affiliation: faculty of engineering Tanta University Egypt

CONTENTS :

Challenges of Wet Briquetting from Locally Available Biomass with Special Reference to Assam

Bichitra Bikash[1], Rajib Bhowmik[2], Madhurjya Saikia[3]

[1]*Mehanical Deptt., Assam Down-Town University, Ghy, Assam*
[2]*Mechanical Deptt., Girijananda Chowdhury Inst. Of Management and Technology, Ghy, Assam*
[3]*Mechanical Deptt., Dibrugarh University Inst. Of Engineering and Technology, Dibrugarh, Assam*

ABSTRACT:
  This study aims at solving energy crisis in rural area via fuel briquettes from locally available biomasses by a well proven technique called wet briquetting. This technique has different operational stages of briquette production. The challenges faced during each operational stage of briquette production are discussed and solutions of the respective problems are tried to be found as well in order to perfect the method. An economic analysis of this method is also done to show profitability margin.

KEYWORDS: *Wet briquetting, biomass, briquetting, economic analysis, durability*

I. INTRODUCTION

With growing development of Indian economy, energy consumption is increasing day by day. Energy consumption in household shares 40% of total energy consumption all over India. Moreover about 30% of total population resides in the villages which consider a good sum of 0.36 billion of total population. In the domestic household sector cooking is the largest end user accounting for almost 90 percent of the total domestic energy use. The rural masses mostly depend on biomass or kerosene for their energy needs. Gradual price hike in crude oil in international market has greatly affected the rural India. In order to cushion fuel price hike, the rural masses are shifting more to biomass. Deforestation for fuel wood has graven the problem of climate change and global warming. The seriousness of the problem can be sensed by seeing depleting forest reserves. This trend needs to be checked from environment point of view. Development of renewable energy sources helps to reduce the degree of dependence on energy imports as well as it can be a tool for curbing carbon emission. So, emphasis is given to the renewable energy program.

The energy requirement in rural household is mainly for cooking and sometimes heating in colder regions. So there is enormous demand for fuel wood. The one option could be the densification or briquetting to counter this problem. It has a great scope in rural India as India produces large amounts of bio waste material every year. This includes rice straw, wheat straw, coconut shells and fibers, rice husks, stalks of legumes and sawdust. Some of this biomass is just burnt in air; some like rice husk are mostly dumped into huge mountains of waste. Open-field burning has been used traditionally to dispose of crop residues and sanitize agricultural fields against pests and diseases. Instead of burning down these wastes or letting to decompose in open air which raises the problem of GHG production, it can be converted to bio fuels to produce power either by direct combustion or transforming these loose biomass to solid fuels [1, 2]. So these processes become automatic candidates for financing under CDM mode [3].

Biomass briquetting is the densification of loose biomass material to produce compact solid composites of different sizes with the application of pressure [4]. Three different types of densification technologies are currently in use. The first, called pyrolizing technology relies on partial pyrolysis of biomass, which is mixed with binder and then made into briquettes by casting and pressing. The second technology is direct extrusion type, where the biomass is dried and directly compacted with high heat and pressure. The last type is called wet briquetting in which decomposition is used in order to breakdown the fibers. On pressing and drying, briquettes are ready for direct burning or gasification. Some of the advantages of briquettes are given below

[1]   This is one of the alternative methods to save the consumption and dependency on fuel wood.
[2]   Densities fuels are easy to handle, transport and store.
[3]   They are uniform in size and quality.
[4]   The process helps to solve the residual disposal problem.
[5]   The process assists the reduction of fuel wood and deforestation.

[6]   Indoor air pollution is minimized.
[7]   Briquettes are cheaper than COAL, OIL or LIGNITE
[8]   There is no sulfur in briquettes.
[9]   There is no fly ash when burning briquettes.
[10] Briquettes have a consistent quality, have high burning efficiency, and are ideally sized for complete combustion.
[11] Combustion is more uniform compared to coal.
[12] Unlike coal, lignite or oil, briquettes are produced from renewable source of energy, biomass.
[13] Loading/unloading and transportation costs are much less and storage requirement is drastically reduced.
[14] Briquettes are clean to handle & can be packed in bags for ease of handling & storage.
[15] Briquettes are usually produced near the consumption centers and supplies do not depend on erratic transport from long distances.
[16] The technology is pollution free and Eco-friendly.
[17] The briquette is easy to ignite.
[18] Continuous burning and long burning duration.

II.   CHALLENGES IN WET BRIQUETTING TECHNOLOGY

The conventional briquetting technologies are capital intensive and unfriendly for smaller scale production. There is a technique called wet briquetting which involves less capital and very low technical machinery which can suit the rural environment for production of briquettes. It is possible to form briquettes from waste crop residues, using a wet process with a hand operated press [5, 6]. First of all suitable biomass is selected. The biomass is decomposed under control environment which is later on pressurized to briquettes. The steps are given below

1) Selection of suitable biomass
2) Decompose biomass
3) Pressurization to form wet briquettes
4) Sun dry wet briquettes

A. Parameters of Selection of biomass

While selecting biomass for wet briquetting, emphasis is given on the local availability of certain type of biomass with lower lignin and ash content. Rice straw, wheat stalks, maize stalks, cotton stalks and barley stalks are some locally available loose biomass or agro residue in rural India. But, the entire available agro residue is not suitable for wet briquetting. For wet briquetting, biomass material is needed to be decomposed before compaction to briquettes. The decomposition period of lignocellulosic biomass depends largely on their lignin content. High lignin containing biomass takes longer time for decomposition. Similarly, biomass having higher ash content is not acceptable for conversion to solid fuel as ash forms clinkers and chances of buildup on the burn pot surfaces, restricting air flow and influencing the removal of ash from the . High ash content also means more frequent dumping of the ash pan. Table 1 shows lignin and ash contents of some locally available agro residues.

Table 1

Lignin and ash contents of some locally available biomasses [7, 8, 9]

Fiber source	Rice straw	Banana fronds	Wheat straw	Barley straw	Maize stalks	Cotton stalks
Lignin (wt %)	9.9	8.0	8.9	13.8	41.0	21.5
Ash (wt %)	17.5	4.7	5.5	10.3	10.2	3.7

B. Parameters governing decomposition of biomass:

Though for other purposes, information on decomposition of lignocellulosic biomass like rice straw is available. Studies on decomposition of brittle rice straw having lower cellulose content revealed that rice straw decomposes fast by anaerobic mechanism when it is incorporated to soil under continuously flooded condition [10]. It is found that at 25ºC under non-flooded conditions, the equivalent of 55% of the rice straw added was mineralized compared to 27% at 58ºC, after 160 days of incubation in soil. Under flooded conditions, the equivalent of 47% of the straw C added was mineralized at 25ºC compared to 19% at 58ºC [11]. The temperature range for optimal decomposition of organic matter is between 52ºC and 60ºCfor aerobic condition [12, 13]. On other hand Acharya et. al., 1935 [14] found that aerobic decomposition of rice straw at about 30ºC

is more than that of anaerobic decomposition. He conducted tests in aerobic, anaerobic and water logged condition on rice straw at 30ºC.

Among all condition, decomposition was highest in aerobic condition within a period of 6 months. But lignin decomposition was found in higher amount in water logged condition in which biomass specimen was kept one inch below water level. The tests show same trend on use of ammonia. Decomposition of biomass feedstock can also be enhanced by application of some fungi or bacteria. Hesham et al. 2006 [15] performed tests on rice straw with actinomycetes and observed a weight loss of 61% within a period of 2months. The high carbon content, high solid content and the low nitrogen content of rice straw require the use of other sources of nitrogen and water to get the proper substrate for the anaerobic digestion process. Nitrogen can be added in inorganic form (ammonia) or in organic form (urea, animal manure or food wastes). Addition of chopped rice straw to dairy manure enhanced the anaerobic digestion process and increased the methane production rate (Hills & Roberts 1981) i.e. more methane means higher carbon mineralization. Somayaji & Khanna et al., 1994 [16] confirmed that addition of chopped rice straw to cattle dung enhanced the organic matter degradation to a high extent (35–51%).

Apart from other biomass, lignocellulosic biomass like rice straw needs some pre-treatment to enhance its degradation. Zhang& Zhang et al., 2006 [17] showed that without thermal pre-treatment, grinding resulted in a significant improvement in terms of solid reduction. Jagdish et al. [18] in his study on wheat straw found that straw size should not increase 1cm. Lower size residue becomes more accessible for the initial microbial attack and led to an enhanced stabilization of microbial biomass. The impact of temperature is immense and it is widely accepted environmental variable. Finstein and Morris, 1975; Finstein et al., 1986 [19] found that minimum temperature level is necessary for high rates of decomposition. MacGregor et al. (1981) [12] found that optimum composting temperatures, based on maximizing decomposition, were in the range of 52–60ºC for aerobic condition. This evidence has supported by their findings (Bach et al., 1984; McKinley and Vestal, 1984) [13]. On the other hand maximum yield in case of anaerobic condition was found at 25ºC.Moisture as variable impacts metabolic and physiological activities of micro organism as it serves as medium for transport of dissolved nutrients [20]. Too much moisture is not desirable as it inhibits the decomposition by making the process anaerobic due to water logging (Schulze et al., 1962; Tiquia et al., 1996).Many investigators have found that 50–60% moisture content is suitable for efficient composting[13,19] . Liang et al.2002 [21] found in his study that 50% is the minimal moisture requirement and even higher decomposing rate can be obtained by having 60-70% moisture. By increasing the moisture content higher temperature requirement can be offset.

C. Factors influencing the final briquette quality during pressurization

It is important to understand the factors that govern compacting. Chaney et. al., 2005,[22] said that some principle factors are the design of die, the method of load application, loading rate , maximum pressure applied, the time for which that pressure is maintained and material characteristics, for example particle size and moisture content. Usually briquetting needs higher amount of pressure for compression. But natural decomposition process can be used to break fibers down and it facilitates bonding [22]. The minimum pressure requirement is about 1Mpa or less. After compression in a die of a hand press, the briquettes relax and try to come to its original shape. It decides the stability and durability. The stable briquettes have less post die expansion [23] found that the relaxation behavior of briquettes mostly depend on the type of residue. For most types of biomass, maximum rate of relaxation occurred after 10 minutes of removal from die followed by a decreasing relaxation for next 2 hours. Chin et al. [23] propose the following relationship between the relaxed densities and applied die pressure:

$$\rho = a \ln P + b$$

where ρ is the relaxed density, P is the compaction pressure and a, b are empirical constants.

Dwell time during compression is decisive for stability of product. Chin et al., 2004 found that with increase in dwell time, maximum length reduction can be obtained as well as smaller post die relaxation. According to Chaney et al. [22], a hold time greater than 40 seconds does not require rigorous control of this variable can yield briquettes of repeating density. Particle size of biomass feedstock is crucial for briquetting. Kaliyan and Morey (2006) [24]indicated that generally, the finer the grind, the higher the quality of compact in case of dry briquetting. Moisture content plays an important role in briquetting of biomass materials. If the moisture content is too high there is a decrease in density and stability. On the other hand, Bellinger et al. [26] showed that energy required to form briquettes is less when there is higher moisture content in the feed stock. Higher moisture in biomass feedstock is desirable for wet briquetting.

Particle size of biomass feedstock is crucial for briquetting. Kaliyan and Morey (2006) [24]indicated that generally, the finer the grind, the higher the quality of compact in case of dry briquetting. In wet briquetting

too same trend is followed. Density of briquettes influences its durability which in turn represents handling characteristics. Durability represents the handling characteristics. Durability for briquettes is measured by following ASAE S269.2 method [25]. According to Saptoadi et al., 2008, briquettes should not be more than 100 g for proper burning [27]. Moreover, a centre hole in briquettes facilitates easier burning of briquettes. Particle size also plays important role in combustion as the voids between particles will be less and less space is available for mass diffusion e.g. water, volatile matter etc [27].

D. Parameters influencing drying of wet briquettes

After the decomposed biomass material are given suitable cylindrical shape with centre hole by a piston press, it should be removed carefully from die and moved aside to dry with minimum handling[25, 28]. It should be ensured that these are placed at a windy place so that briquettes could get even air flow across their whole surface. Fuel briquettes generally take three to six days to dry direct sunlight and in cloudy condition it may increase up to eight to then days.

III. ECONOMIC VIABILITY OF WET BRIQUETTING

Before setting up any enterprise, the cost benefit analysis is a must. Then we can forecast the profitability of the enterprise. Sometimes a project even if it looks good may turn out to be fruitless in terms of economic analysis. To be a successful project, it must overcome economic barrier. Therefore, to understand fully the financial feasibility of briquette production, we need to do some preliminary feasibility exercise. The main objective of economic analysis is to compare the cost of briquette production per day per family to cost f fuel wood usage per day per family [28].

Table 1: Parameters for economic analysis

Parameters	Values
Daily wood requirement for a family of 4 members	7
Cost of wood per kg(Taking average), Rs	5
Worker cost per day ,Rs	150
Requirement of worker for the project	6
Maintenance and equipment cost added to worker cost,%	15
Briquettes used per family at an average	12-15

Daily fuel wood cost for the family= Rs 35 per day
Labour cost per day per person=Rs 150
Total labour cost for 6 labors = Rs 900 which will be producing briquettes between 750-1000 briquettes.
Adding 15% for other minor costs, such as equipment and maintenance = Rs 1035.
This is then divided by the 50 families who would be served by this production to arrive at a daily cost of fuel briquettes of Rs 20.7 per day per family.
Therefore, in comparison to the Rs 35 a day for fuel wood, the cost of making fuel briquettes at Rs 20.7 provides a feasible margin for the group to begin briquette production.
Moreover, the cost of production of each briquette can be determined by

$$\text{Cost of briquette} = \frac{6 \times \text{Average Daily Labour cost}}{\text{Average daily Production of 1000 briquettes}}$$

So, in this case the cost of each briquette comes around Rs 0.9. Also it has been noticed from extensive studies an ideal family having 4 to 5 members uses 12-15 briquettes per day. So, total fuel cost of a family is around Rs 18 which is about Rs 17 cheap than that of wood.

IV. RESULTS AND DISCUSSIONS

Wet briquetting depends on the decomposition of biomass materials such as various crop residues. From the above studies we come to opinion that decomposition of finely chopped biomass at anaerobic condition is faster. Moreover, by keeping biomass materials in heap condition at sun will enhance decomposition. However, during compaction of briquettes, wet biomasses need to be kept on pressing at least for 40 seconds and compaction pressure should not be less than 1 Mpa for the purpose to yield good quality briquette. During drying of briquettes, wet briquettes should be placed at windy places so that air circulates around its surfaces. The studies also indicate that briquette should be dried up to 8% moisture content otherwise it will cause severe smoke formation during burning. It has also come to notice that briquettes weighing above

100 gm shows problem during burning and handling. Therefore optimum weight should be less or equal to 100 gm. A cylindrical shaped briquette with a central hole burns at ease .

V. REFERENCES

[1]   Campbell C A; Zentner R P; Gameda S; Blomert B; Wall D D (2002). Production of annual crops on the Canadian prairies: trends during 1976–1998. Canadian Journal of Soil Science, 82, pg 45–57.
[2]   Mania S., Tabil L.G. and Sokhansanj S. (2006). Effects of compressive force, particle size and moisture content on mechanical properties of biomass pellets from grasses, Biomass and Bioenergy, Vol 30, pg 648–654. [3] Kishore et al., 2004). Biomass energy technologies for rural infrastructure and village power-opportunities and challenges in the context of global climate change concerns, Energy Policy, Vol 32, pg 801-810.
[4]   Grover P.D. and Mishra S.K. (1994). Development of an appropriate biomass briquetting technology suitable for production and use in developing countries Energy for Sustainable Development, Vol 1.
[5]   Chaney J. O., Clifford M. J., Wilson R. An experimental study of the combustion characteristics of low-density biomass briquettes. Biomass magazine 2010, Vol 1.
[6]   Stanley, R. Fuel Briquettes-Theory and applications from around the world, Legacy Foundation, (2003).
[7]   Stanley, R. Fuel Briquettes-Theory and applications from around the world, Legacy Foundation, (2003).
[8]   Zohar k. And Hadar Y. (1995).Effect of Manganese on Preferential Degradation of Lignin by *Pleurotus ostreatus* during Solid-State Fermentation, Applied And Environmental Microbiology, Vol 65, pg 3057–3062.
[9]   Han Y. W. and Lee J.S. (1975).Chemical Composition and Digestibility of Ryegrass Straw, J. Agric. Food Chem., Vol. 23, pg 931.
[10]   Johnson Sarah, Brar D.S. and Buresh R.J. (2006). Faster anaerobic decomposition of a brittle straw rice mutant: Implications for residue management, Soil Biology & Biochemistry, vol 38, pg 1880–1892.
[11]   Oliver et al., 2000. MNRAS, 316, 749.
[12]   MacGregor, S.T., Miller, F.C., Psarianos, K.M., Finstein, M.S., (1981). Composting process control based on interaction between microbial heat output and temperature. Appl. Environ. Microbiol. 41, 1321-1330.
[13]   McKinley, V.L., Vestal, J.R., Eralp, A.E. 1986. Microbial activity in composting. In: The biocycle guide to in-vessel composting. JG Press Inc., Emmaus, Pennsylvania, pp. 171–181
[14]   Acharya C.N (1933). Studies on the anaerobic decomposition of plant materials, Biochem.J, Vol 29, pg528.
[15]   Hesham et al. 2006. Investigation of optimum conditions of co-composting process by using of sewage sludge and municipally waste, The 1th International and the 4th National Congress on Recycling of Organic Waste in Agriculture, Iran.
[16]   Somayaji, D. & Khanna, S. 1994. Biomethanation of rice and wheat straw. W. J.Microbiol. Biotechnol. 10: 521–523.
[17]   Zhang& Zhang et al., 2006. Efficient conversion of wheat straw wastes into bio hydrogen gas by cow dung compost, Bioresource Technology, Volume 97, Issue 3 Pages 500–505.
[18]   Gabhane, Jagdish., William, S.P.M. Prince., Vaidya, Atul Narayan., Mahapatra, Kalyani.,& Chakrabarti, Tapan., Influence of heating source on the efficacy of lignocellulosic pretreatment – A cellulosic ethanol perspective, Biomass and Bioenergy, 35 (1), 96-102, 2011.
[19]   Finstein M.S., Morris M.L. (1975). Microbiology of municipal solid waste composting. Adv. Appl. Microbiol., 19: 113-151.
[20]   McCartney, D., Tingley, J. (1998). "Development of a rapid moisture content method for compost materials". Compost Sci. 6, 14-25.
[21]   Liang C and McClendon R. W. (2003).The influence of temperature and moisture contents regimes on the aerobic microbial activity of a bio solids composting blend, Bio resource Technology, vol 86, pg 131–137.
[22]   Chaney J. O., Clifford M. J., Wilson R. An experimental study of the combustion characteristics of low-density biomass briquettes. Biomass magazine 2010, Vol 1.
[23]   Chin Ooi and Siddiqui K.M. (2000). Characteristics of some biomass briquettes prepared under modest appliied die pressures, Biomass and Bioenergy, Vol 18, pg 223 to 228.
[24]   Kaliyan N. and Morey R.V. (2008). Factors affecting strength and durability of densified of biomass products, Biomass and Bioenergy, Vol 33, pg 337-359.
[25]   Saikia M and Baruah D. (2013). Analysis of Physical Properties of Biomass Briquettes Prepared by Wet Briquetting Method, International Journal of Engineering Research and Development, Volume 6, Issue 5 (March 2013), PP. 12-14.
[26]   Bellinger, P.L. and H.H.McColly, Energy requirements for forming hay pellets, Journal of Agricultural Engineering, 42, 1961, 244-247.
[27]   Saptoadi, H., 2008. The best biobriquette dimension and its particle size. Asian J. Energy Environ., 9: 161-175.
[28]   Stanley R. (2003).Fuel Briquette making, Legacy Foundation.
[29]   Saikia M, Bhowmik R, Baruah D, Dutta B and Baruah D.C. (2013). Prospect of bioenergy substitution in tea industries of North East India**,** International Journal of Modern Engineering Research (IJMER) Vol.3, Issue.3, May-June. 2013 pp-1272-1278.

Neuromuscular Activities On Lower Limb's Joint Contact Forces During Normal Human Walking

Biswajit Bera
Department of Mechanical Engineering, NIT, Durgapur, India

ABSTRACT:

Present study describes neuromuscular activities on lower limb's joint contact forces during normal human walking. First of all, a biomechanical human model has developed to evaluate accurate lower limb's joint contact forces during normal human walking. According to author postulation, lower limb's joint contact forces should not be greater than the ground contact force (ground reaction). It is found that maximum total contact forces of ankle joint, knee joint and hip joint during normal human walking are approximately, 100%, 95% and 85% of body weight of the subject respectively. Total joint contact force is compressive and tensile for respective stance phase and swing phase of human walking gait cycle. Finally, the neuromuscular activities on the joint's contact forces clearly explained on the basis of EMG response of the subject.

KEYWORDS: *Biomechanical human model, Joint contact forces, Neuromuscular activities*

I.   INTRODUCTION

Musculoskeletal system produces very important role to perform different human movements. Among them, walking is very common and frequent human movement. Still today, the walking mechanism is not clearly, understood because knowledge of musculoskeletal loading of lower limb's joint during human walking is limited. In this field, most of the scientists used optimization technique to evaluate lower limb's joint contact forces during human walking. Joint contact forces are generated by combined effect of muscle forces across a joint during normal human walking. Theoretically, there are two category of optimization method; one is static optimization and other is dynamic optimization. The static optimization method has been used extensively to estimate in vivo muscle forces [1-6]. Also, The dynamic optimization method is used to find out in vivo muscle forces [7-9]. On the other hand, experimentally, the joint contact forces are also measured by some pioneering scientists [10-14]. Both of the two way, theoretically and experimentally, still today, it is not possible to find out the accurate value of lower limb's joint contact forces during normal human walking. Still now, It is reported that range of calculated joint contact forces are within 200% to 600% of body weight during normal human walking whereas boundary value of joint contact force is the ground contact force (ground reaction) [13]. It is well known that the highest range of ground contact force is within 120% of body weight [15]. According to author's postulation, lower limb's joint contact forces should not be greater than the ground contact force during normal human walking. So, contact forces could be evaluated accurately from ground contact force considering dynamic equilibrium of lower limb's joint. Thereafter, neuromuscular activities on total contact force of the lower limb's joint would be explained on the basis of EMG response for a particular subject (person).

II.   THEORETICAL FORMULATION

2.1 Biomechanical human model

Human body is modeled as a 3D system of seven segments of articulated, rigid massy linkage with 8 degree of freedom as shown in schematic diagram Fig.1. Here, head, arms, torso and pelvis are represented as a single rigid body, trunk. The remaining 6 segments are branched out in two parts from respective two hip center

and each branch part could be considered as a mechanical chain of articulated three lower limbs, thigh, shank and foot. All three joint of lower limbs have taken as a perfect hinge joint (DOF 1) to satisfy the walking in
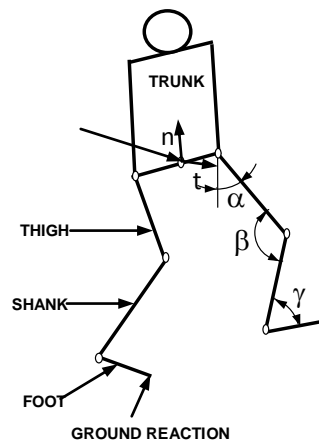
Fig. 1 Diagram of human model

sagital plane only. According to model, it is assumed approximately that C.G. of the human body is situated on middle of the line joining two hip centers. The intrinsic coordinate system is fixed at C.G. of human body as shown in Fig.1. The t-axis is directed forward, tangential to ground and n-axis is directed upward normal to ground. More specifically, it should be mentioned that the n-t coordinate system is selected on the basis of fundamental mechanism of human walking. As gravity force of body weight acts towards the center of earth normal to the earth surface (n-dir^n), so, human walking is a natural balanced process for shifting the gravity force of body weight tangential to the earth surface (t-dir^n). The 8 DOF systems consists of two DOF of center of gravity (CG) for curvilinear motion along n and t direction, and single DOF of hip, knee, and ankle joints for the angular displacement. α, β and γ represents relative angular coordinates of thigh, shank and foot respectively.

2.2 Acceleration of lower limbs
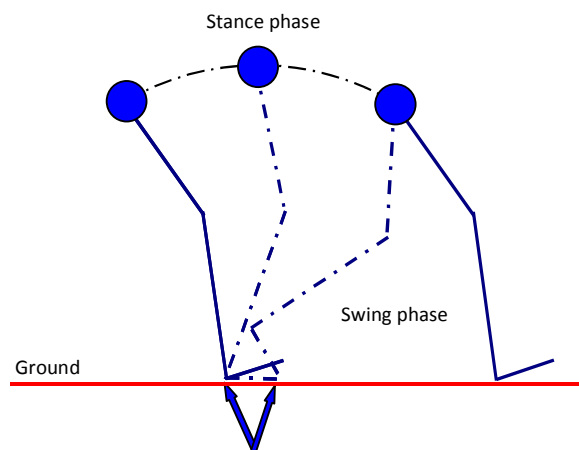
Fig.2 Line diagram of walking simulation

Human walking is a very complex balanced motion adopted from childhood naturally. It is an alternation of stance phase and swing phase cyclically for propagation of CG of human body toward forward direction as shown in Fig 2. During human walking foot behave like a wheel segment. So, the stance phase should be considered as clockwise rolling of CG over foot wheel with pendulum motion of lower limbs. During rolling of whole body, accelerations of lower limbs change gradually, from foot to shank and shank to thigh

causing acceleration of CG in curvilinear path. Considering, over all free body diagram of human body, following equations are obtained along respective tangential and normal direction [16].

Tangential acceleration of CG = (Tangential ground reaction / body weight)* g
Normal acceleration of CG   =   (Normal ground reaction / body weight – 1)* g

According to assumption of modeling, as CG lies over mid point of the line joining two hip centers, so, acceleration of hip center equals to the acceleration of CG. Thereafter, components of acceleration of lower limbs could be determined in terms of generalized coordinates of thigh, shank and foot. Generalized coordinates (q_i) of respective lower limbs are expressed in term of known value of relative angular coordinate of thigh (α), shank (β) and foot (γ) with respect to vertical axis parallel to normal (n) axis through the following set of eq[n] 1.

$$
\begin{cases}
\text{Generalized coordinate of thigh}, & q_1 = \alpha \\
\text{Generalized coordinate of shank}, & q_2 = \alpha + \beta - \pi \\
\text{Generalized coordinate of foot}, & q_3 = \alpha + \beta - \gamma
\end{cases}
\qquad
\begin{aligned}
& q_1 = 0.569 - 3.915t + 3.857t^2 + 6.286t^3 \\
& q_2 = 0.492 - 5.977t - 0.156t^2 + 17.210t^3 \\
& q_3 = 1.80 - 5.524t + 0.0935t^2 + 14.884t^3
\end{aligned}
\tag{1}
$$

As during stance phase, the lower limbs are not only rotating with the rotation of CG of whole body but also,
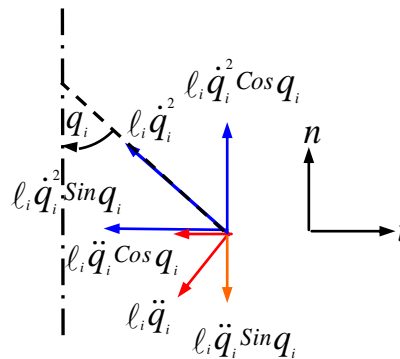
Fig. 3 Schematic accelerations of a limb

$$
\begin{cases}
\text{Tangential acceleration}, & a_i^t = -\ell_i \ddot{q}_i Cosq_i - \ell_i \dot{q}_i^2 Sinq_i \\
\text{Normal acceleration}, & a_i^n = -\ell_i \ddot{q}_i Sinq_i + \ell_i \dot{q}_i^2 Cosq_i
\end{cases}
$$

they follow the three limb pendulum motion, so, q_1, q_2, and q_3 will not change equally. During swing phase, three lower limbs completely, follow the three limbs pendulum motion about human CG. As CG of whole body rotates clockwise direction, the CG of lower limbs also rotates clockwise direction with respect to the CG of whole body. Generalized expression for components of a lower limb's acceleration is evaluated from the algebraic sum of components of centripetal acceleration ($\ell_i \dot{q}_i^2$) and tangential acceleration ($\ell_i \ddot{q}_i$) of the limb along the respective tangential and normal direction as shown in Fig.3. Now, components of acceleration of thigh relative to hip joint, these of shank relative to knee joint and these of foot relative to ankle joint are evaluated for respective proximal length of thigh (ℓ_t^p), shank (ℓ_s^p) and foot (ℓ_f^p) as shown in eq[n] 2, 3, 4.

Acceleration of thigh

$$
\begin{cases}
a_{thigh}^t = a_{hip}^t - \ell_t^p \ddot{q}_1 Cosq_1 - \ell_t^p \dot{q}_1^2 Sinq_1 \\
a_{thigh}^n = a_{hip}^n - \ell_t^p \ddot{q}_1 Sinq_1 + \ell_t^p \dot{q}_1^2 Cosq_1
\end{cases}
\tag{2}
$$

Acceleration of shank

$$\begin{cases} a^t_{shank} = \underbrace{a^t_{hip} - \ell_t\ddot{q}_1 Cosq_1 - \ell_t\dot{q}_1^2 Sinq_1}_{acc^t \ at \ knee \ joint} - \ell_s^p\ddot{q}_2 Cosq_2 - \ell_s^p\dot{q}_2^2 Sinq_2 \\[2em] a^n_{shank} = \underbrace{a^n_{hip} - \ell_t\ddot{q}_1 Sinq_1 + \ell_t\dot{q}_1^2 Cosq_1}_{acc^n \ at \ knee \ joint} - \ell_s^p\ddot{q}_2 Sinq_2 + \ell_s^p\dot{q}_2^2 Cosq_2 \end{cases} \tag{3}$$

Acceleration of foot

$$\begin{cases} a^t_{foot} = \underbrace{a^t_{hip} - \ell_t\ddot{q}_1 Cosq_1 - \ell_t\dot{q}_1^2 Sinq_1 - \ell_s\ddot{q}_2 Cosq_2 - \ell_s\dot{q}_2^2 Sinq_2}_{acc^t \ at \ ankle \ joint} - \ell_f^p\ddot{q}_3 Cosq_3 - \ell_f^p\dot{q}_3^2 Sinq_3 \\[2em] a^n_{foot} = \underbrace{a^n_{hip} - \ell_t\ddot{q}_1 Sinq_1 + \ell_t\dot{q}_1^2 Cosq_1 - \ell_s\ddot{q}_2 Sinq_2 + \ell_s\dot{q}_2^2 Cosq_2}_{acc^n \ at \ ankle \ joint} - \ell_f^p\ddot{q}_3 Sinq_3 + \ell_f^p\dot{q}_3^2 Cosq_3 \end{cases} \tag{4}$$

2.3 Dynamic equilibrium of lower limbs
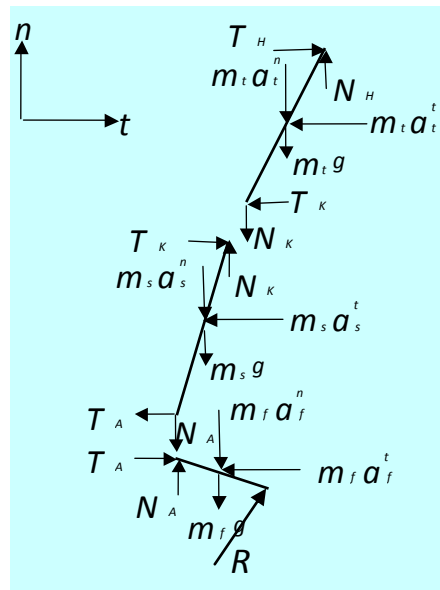
Fig.4 Force equilibrium of lower limbs

According to biomechanical model, human body is considered as seven segments of articulated, rigid massy linkage. Lower limbs consist of foot, shank, and thigh. Free body diagram of each limb is represented by gravity force, joint contact force and inertia force as shown in Fig.4. It should be mentioned that joint moment is not shown in the free body diagram of each limb to avoid complexity of the diagram. Each lower limb must maintain the dynamic equilibrium during human locomotion. Applying Newton second law of motion for each limb along tangential and normal direction, the following equation of motion for each limb can be written.

$$\begin{cases} \sum F^t = ma^t \\ \sum F^n = ma^n \end{cases}$$

Contact force equations of lower limb's joint (eq[n] 5, 6, 7) could be obtained from respective force equilibrium of lower limbs, foot, shank, and thigh.

Contact force equation of lower limb's joint

Contact force equations of ankle joint

$$\begin{cases} T_A = -[T_R - m_f a_f^t] \\ N_A = -[N_R - m_f g - m_f a_f^n] \end{cases} \tag{5}$$

Contact force equations of knee joint

$$\begin{cases} T_K = -[T_R - (m_f a_f^t + m_s a_s^t)] \\ N_K = -[N_R - (m_f g + m_s g) - (m_f a_f^n + m_s a_s^n)] \end{cases} \tag{6}$$

Contact force equations of hip joint

$$\begin{cases} T_H = -[T_R - (m_f a_f^t + m_s a_s^t + m_t a_t^t)] \\ N_H = -[N_R - (m_f g + m_s g + m_t g) - (m_f a_f^n + m_s a_s^n + m_t a_t^n)] \end{cases} \tag{7}$$

III. RESULTS AND DISCUSSION

For input data (Appendix A), **subject B of A Pedotti work** is considered as a standard patient [15]. Test parameters and body parameters are taken for the subject B. Similarly, ground reaction and relative angular displacement are taken for the subject B also. Through synchronization in between ground reaction and kinematics variable, a film was taken by movie camera whose optical axis was orthogonal to the direction of propagation, during the stride on the force plate. By this way, relative angular displacements of lower limb's joints were measured [15]. Generalized angular displacements are obtained through the set eq[n]1 for the respective lower limb's joints. Thereafter, the generalized coordinates (q_1, q_2, q_3) are cubic spline curve fitted by least square method (eq[n]1).
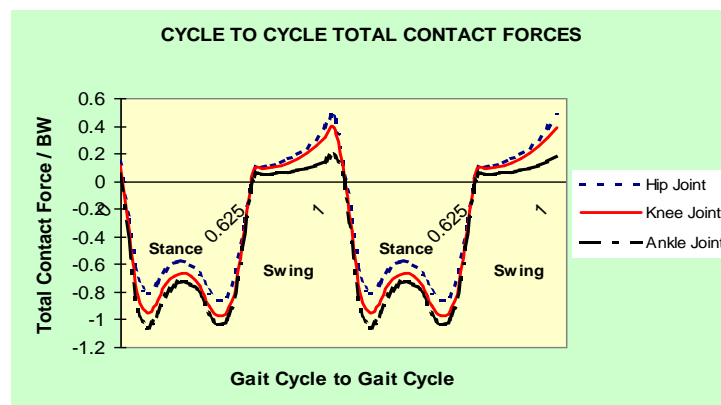
Fig. 5 Cycle to cycle total contact forces

Fig.5, depict cycle to cycle total contact forces for ankle joint, knee joint and hip joint respectively. Total contact force is compressive and tensile for respective stance phase and swing phase of walking gait cycle of the subject. It is found that maximum total contact force of ankle joint, knee joint and hip joint are approximately, 100%, 95% and 85% of body weight of the subject respectively. Gradually, decrement of total contact force of ankle joint to knee joint and knee joint to hip joint occurs due to mainly respective gradual decrement of normal contact forces. Gradually, the normal contact forces are decreases from ankle joint to hip joint due to gradual subtraction of gravity forces and inertia forces of lower limbs that is clearly, stated in normal contact forces equations. Actually, compressive and tensile total contact forces are generated by muscle forces during normal human walking. The neuromuscular activities of lower limb's joint contact forces during stance phase and swing phase could be explained on the basis of EMG response of each muscle. Human walking simulation (Fig.2) is drawn on the basis of relative angular displacement for he subject B approximately to understand the neuromuscular muscular activities on joint contact force during human walking gait cycle. EMG response is also considered for the subject B as a standard patient [15]. In the discussion stance phase is considered in two parts.

- **Stance phase**
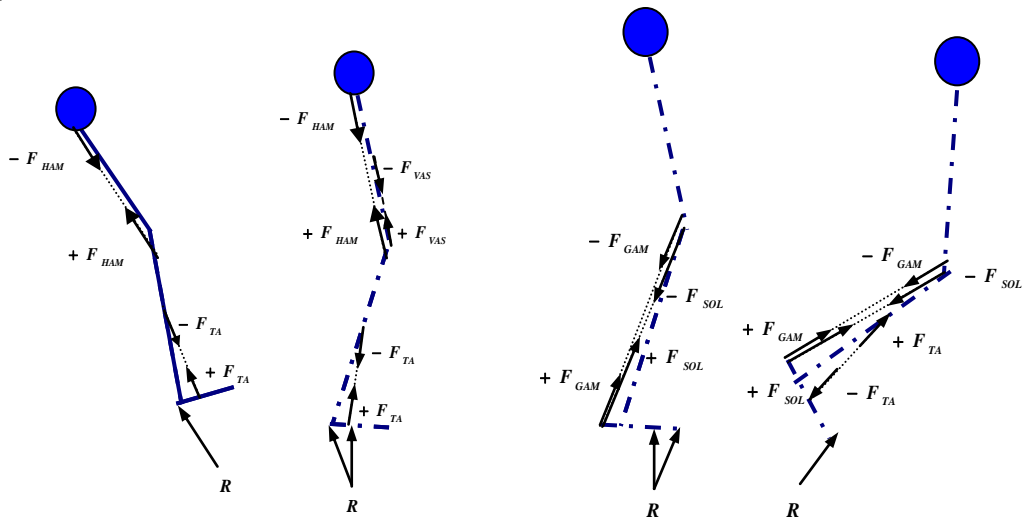
Fig. 6 Early stance phase

Fig. 7 Late stance phase

Earlystance phase

From the EMG response of the subject B, it is found that Tibialis interior (TA), Semitendinosous and Semimembranosus (Hamstring, HA) muscle are active during heel strike. The TA produces compressive contact force across the ankle joint whereas HA produces compressive contact force across the knee joint and hip joint as shown in Fig.6. As body rolls over heel wheel segment, muscle action of HA and TA increases by decreasing relative angular displacement of ankle joint and knee joint and hip joint. It results the increment of linear loading of compressive total contact force. It should be mentioned that the compressive total contact force is maximum just after heel strike causing first peak of total contact force as shown in Fig.5. After heel strike, the Hamstring muscle (HA) relaxes gradually, but muscle action of Vastus medialis and Vastus lateralis (VAS) starts action on the tribofemoral knee joint (Fig.6) and still, compressive action of Tibialis interior continues further decrement of the dorsiflexion angle of ankle joint. Compressive action of VAS produces gradual decrement of compressive contact force due to stretching of three lower limb pendulum as shown by convex profile in Fig.5.

Late stance phase

Thereafter, from the EMG response, it is found that TA and VAS relaxes and Gastrocnemius (GA) and Soleus (SOL) produces compressive muscle force which causes toe strike on the ground and starts rolling of whole body over the toe wheel segment. Both of the compressive muscle GA and SOL generates compressive contact force across the ankle joint whereas only GA produces similar nature of contact forces across the knee joint and hip joint (Fig.7). Action of these muscle force increases again upto maximum value just after end of toe strike which causes second peak value of total contact force as shown in Fig.5. During toe off, Tibialis interior (TA) muscle exerts gradually, very high tensile force for lifting foot from the ground as shown in Fig.7. The tensile TA muscle force gradually decreases compressive contact force linearly (upto zero i. e. toe off) as shown in Fig.5.

Swing Phase

.
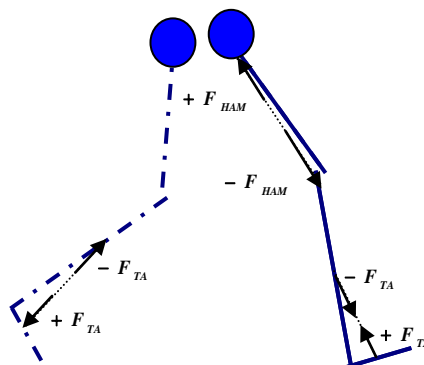
Fig.8 Swing phase

After toe off, from the EMG response it is found that the compressive Gastrocnemius (GA) and Soleus (SA) do not produce any muscle force and only, Tibialis interior (TA) becomes active. The TA produces maximum plantar flexion of ankle joint and flexion of knee joint and hip joint by making almost like ' **Z** ' shape of three lower limbs just after toe off as shown in Fig.8. Thereafter, Hamstring (HA) muscle should produce tensile action providing swing of foot and shank about the knee joint. Possibly, the tensile HA muscle force produces tensile nature of contact force across the lower limb's joint during swing phase. It should be mentioned that TA changes action from tensile to compressive at the end of swing phase for preparation of heel strike that causes starting of next stance phase.

IV. CONCLUSION

The aim of this study was to explain the neuromuscular activities on lower limb's joint contact forces during normal human walking. Without optimization technique, joint contact forces have calculated accurately considering dynamic equilibrium of lower limb's joints. It is found that maximum contact force of ankle joint, knee joint and hip joint are approximately, 100%, 95% and 850% of body weight of the subject respectively. From the EMG response of the subject, neuromuscular activities on joint contact forces as summarized below. HAM and TA develop compressive contact force to overcome ground reaction during heel strike. During mid stance phase VAS produces stretching of three lower limbs which causes slight decrement of compressive contact force of lower limb joint. Thereafter, GA, SOL, TA produces joint contact force for toe off i.e. lifting foot from the ground. During swing phase, initially, the three lower limbs are formed like Z shape by tensile action of TA and thereafter, tensile action of HA provides swing of shank and foot with respect to knee joint for smooth heel strike of foot.

REFERENCES

[1]   A. Seireg, R. J. Arvikar, "The prediction of muscular load sharing and joint forces in the lower extremities during walking", Journal of Biomechanics, vol. 8, 1975, pp. 89-102

[2]   A. Pedotti, V. V. Krishnan, L. Stark, "Optimization of muscle force sequencing in human locomotion", Mathematical Biosciences, Vol. 38, 1978, pp. 57-76.

[3]   R. D. Crowninshield, R. C. Johnston, J. G. Andrews, R. A. Brand, "A biomechanical investigation of the human hip", Journal of Biomechanics, Vol. 11, 1978, pp 75-85.

[4]   R. D. Crowninshield, R. A. Brand, "A physiologically based criterion of muscle force prediction in locomotion", Journal of Biomechanics, Vol. 14, 1981, pp 793-801.

[5]   R. A. Brand, D. R. Pedersen, J. A. Friederich, "The sensitivity of muscle force predictions to changes in physiologic cross-sectional area", Journal of Biomechanics, vol. 19, 1986, pp. 589-596.

[6]   D. R. Pedersen, R. A. Brand, D. T. Davy, "Pelvic muscle and acetabular contact forces during gait", Journal of Biomechanics, Vol. 30, 1997, pp. 959-965.

[7]   D. T. Davy, M. L. Audu, "A dynamic optimization technique for predicting muscle forces in the swing phase of gait", Journal of Biomechanics, Vol. 20, 1987, pp. 187-201.

[8]   G. T. Yamaguchi, F. E. Zajac, "Restoring unassisted natural gait to paraplegics via functional neuromuscular stimulation: a computer simulation study", IEEE Transactions on Biomedical Engineering, Vol. 37, 1990, pp. 886-902.

[9]   F. C. Anderson, M. G. Pandy, "Static and dynamic optimization solutions for gait are practically equivalent", Journal of Biomechanics, Vol. 34, 2001, pp. 153-161.

[10]   T. A. English, M. Kilvington, "In vivo records of hip loads using a femoral implant with telemetric output", Journal of Biomedical Engineering, Vol. 1, 1979, pp. 111–115.

[11]   D. T. Davy, G. M. Kotzar, R. H. Brown, K. G. sen. Heiple, V. M. Goldberg, Jr. K. G. Heiple, J. Berilla, A. H. Burstein, "Telemetric force measurements across the hip after total arthroplasty", Journal of Bone and Joint Surgery, Vol. 70-A, 1988, pp. 45–50.

[12]   R. A. Brand, D. R. Pedersen, D. T. Davy, K. G. Heiple, V. M. Goldberg, "Comparison of hip force calculations and measurements in the same patient", Transaction of ORS, Vol. 1, 1989, pp. 96-99.

[13]   G. Bergmann, F. Graichen, A. Rohlmann, "Hip joint forces during walking and running, measured in two patients", Journal of Biomechanics, Vol. 26, 1993, pp. 969–990.

[14]   G. Bergmann, G. Deuretzbacher, M. Heller, F. Graichen, A. Rohlmann, J. Strauss, G. N. Duda, "Hip contact forces and gait patterns from routine activities", Journal of Biomechanics, Vol. 34, 2001, pp 859-871.

[15]   A. Pedotti, "A study of motor coordination and neuromuscular activities in human locomotion", Biological Cybernetics, Vol. 26, 1977, pp. 53-62.

[16]   A. Crowe, P. Schiereck, R. W. de Boer, W. Keessen, "Characterization of human gait by means of body center of mass oscillations derived from ground reaction forces", IEEE Transactions on Biomedical Engineering, Vol. 42, 1995, pp. 293-302.

Appendix A: Input Data

Test Parameters						
Subject	Age	H (m)	W (Kg)	Steps/min	Length (m)	Speed (Km/hr)
Subject B	26	1.80	70.0	104	0.83	5.17

Body Parameters				
Subject B	Length (m)	Weight (kg)	Inertial moment on centre of gravity (kg m^2)	Distance of CG from proximal joint (m)
Thigh	0.40	7.31	0.062	0.164
Shank	0.47	3.22	0.052	0.185
Foot	0.27	1.42	0.001	0.038

Ground Reactions	Angular Displacements

Appendix B: Cycle to cycle contact forces

CYCLE TO CYCLE NORMAL CONTACT FORCES

CYCLE TO CYCLE TANGENTIAL CONTACT FORCES

CYCLE TO CYCLE TOTAL CONTACT FORCES
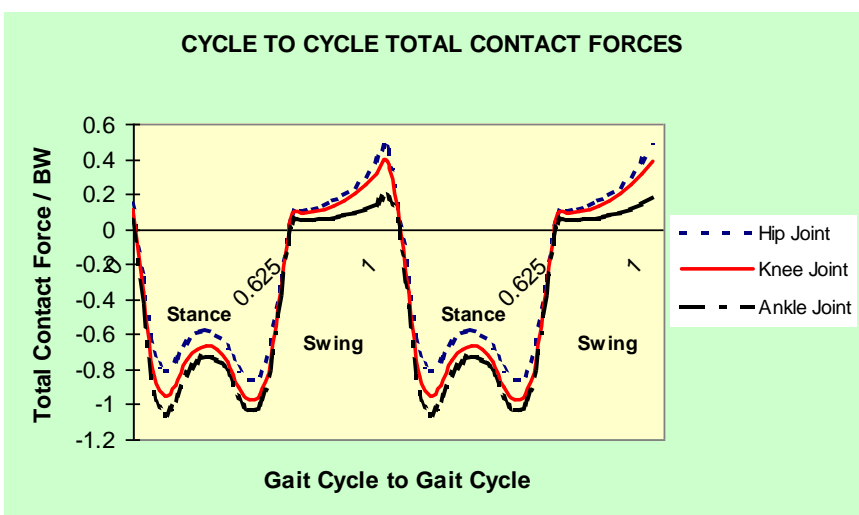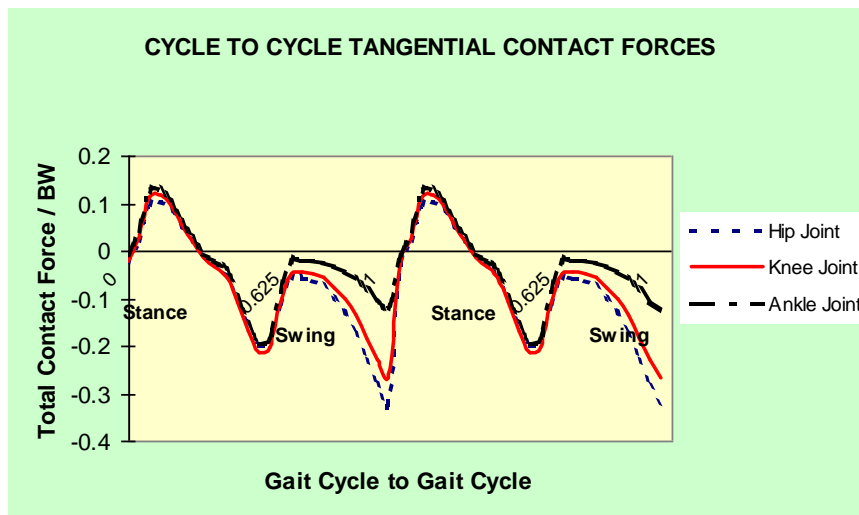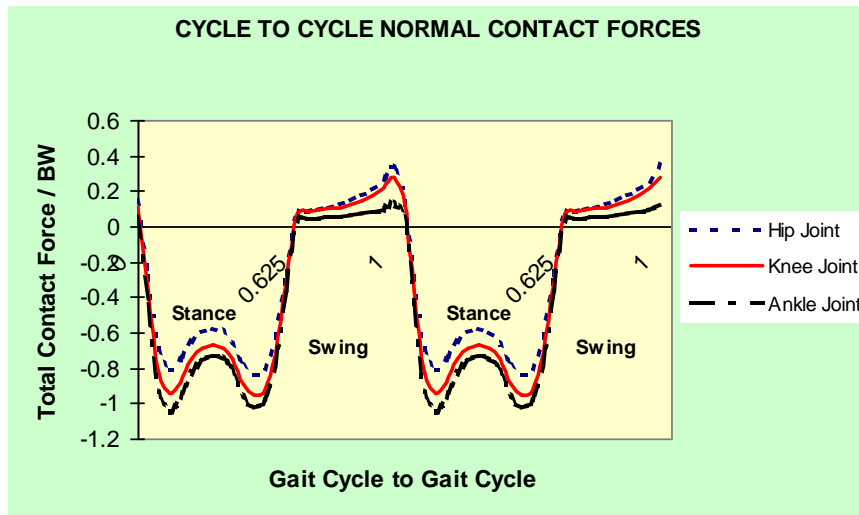
Review Paper: To Reduce The Handover Delay In Wimax When The Mobile Station Moves At Higher Speed.

[1,]Mrs Sukhpreet Kaur, [2,]Jasbir Kaur

ABSTRACT

The real potential of the broadband wireless networks lies with the mobility. WIMAX is a Wireless Interoperability for Microwave Access. It is a telecommunication technology that provides wireless the data over long distances in several ways, from point-to-point links to the full mobile cellular type access. The main consideration of Mobile Wimax is to achieve seamless handover such that there is no loss of data. In this review paper we have compared different techniques which provides no data losses when the mobile station moves at higher speed. In this we discuss the combination of adaptive forward error correction (FEC) with retransmission which provides the extra protection for handover signaling messages to enhance the probability of a successful handover, especially at a higher velocity, Media Independent Handover (MIH) standard which proposes a new neighbor discovery mechanism, considering the hierarchical view of the network information, soft handover to avoid the data losses during handover by using base station selection procedure that will optimize the soft handover such that there is no data loss; handover decision is taken quickly and thus improving overall handover performance.

I.   INTRODUCTION

The introduction of broadband wireless WiMAX solution based on IEEE 802.16 technology makes it possible a standard based low cost solution for the last mile [7]. In particular, with its   coverage of 30 miles and non line of sight technology based on OFDM, it will be able to construct a metropolitan network where broadband access from anywhere within the area is possible. Handover is one of essential issues in mobile wirelessCommunications since it is needed to maintain uninterrupted services during user's movement from one location to another [3], [4]. It manages mobility between subnets in the same network domain (micro-mobility) and between two different network domains. The current mobile WiMAX standard defines handover operations to support micro-mobility for the point-to-multipoint (PMP) mode communication.Most previous work considered the handover performancefrom the viewpoint of the MAC or IP layer, but paid littleattention to the physical layer. That is, physical channelquality and/or user's mobility are not taken into account.Usually, handover occurs in the boundary of BS's coveragearea, where signals transmitted on the radio channel are weakand unstable. When channel quality degrades, both signalling messages and data packets are at a higher risk of being lostor corrupted.Moreover, user's mobility plays an important role in the handover process. The radio channel quality for mobile users at a higher velocity often suffers from severe degradation due to the Doppler frequency shift. Since the overlap area is limited between adjacent BSs, the requirement on handover latency is more stringent for users with higher mobility. In order to support full mobility (which means seamingless handover for users moving at a speed of 120 kilometer per hour or higher), these issues become severer and have to be addressed.

Specification known as IEEE 802.16e-2004 which is wireless but fixed, it lacks the ability for user to move during data transmission. The main purpose of Wimax is to provide users in rural areas with high speed communications as an alternative to expensive wired connections (e.g. cable or DSL). That is Wimax is capable to provide high speed internet to last mile connections. But this is not the only purpose of Wimax systems. Mobile Wimax allows the user to move freely during data transmission. The main consideration of mobile Wimax is that there should be no data loss when the moving user switches from one base station to another i.e. during handover. Handover is procedure when a mobile station changes the serving base station. The reason for handover could be relatively low signal strength or work load of base station.[3] Wimax is a state-of-the-art wireless technology which utilizes adaptive modulation and coding, supports single carrier (SC) and orthogonal frequency division multiplexing techniques (OFDM) and several frequency bands for different operation environments. WiMAX system is able to constantly monitor the quality of the radio channel and change its operational parameters (e.g. modulation and coding) accordingly.

The integration of various types of wireless technologies IEEE 802.11 Wi-Fi and IEEE 802.16 WiMAX allows mobile users to choose an optimum network interface in mobile users to choose an optimum network interface in accordance with the desired requirements in terms of quality of service,price, transmission rate, security and other characteristics. The IEEE 802.21 standard [1] aims to facilitate handover procedure in heterogeneous access networks by providing information, events and commands to the entities that assist in the handover decision. In this heterogeneity of the technologies, discovering the available access network is one of the main challenges. The standard [1] specifies a Media Independent Information Service (MIIS) server supporting various information elements that provide network information within a geographical area. Based on information from several access networks and operators, a mobile node

(MN) can take an optimized handover decision. There have been proposals [2]-[5] taking into consideration the IEEE 802.21 MIIS service for network discovery. Current literature considers the existence of only one MIIS server in the network which responds with neighborhood information. However, the number of network to the MN, causing handover performance delay.

There are many shortcomings related to the specification of single MIIS server
1) Too much information to store
2) Can represent a single point of failure
3) High discovery delay if the MIIS server is located many hops away from the MN.

II.   FEC WITH RETRANSMISSION
To increase the successful handover probability of an MS at a vehicular speed, we need to establish more reliable wireless links in transmitting signaling messages besides using auto retransmission scheme. In this section, we investigate the improvement of the probability of a successful handover by adding extra protection for handover signaling messages. Forward error correction codes provide an error control technique for data transmission, where a sender adds redundant data to messages. These redundant data can assist the receiver detecting and correcting errors to further improve wireless channel quality. During the WiMAX handover procedure, signaling messages mostly consist of small data packets whose size is less than 50 bytes. Consequently, the amount of parity-check data needed for error correction in each message is small and the resource required by error correction does not increase excessively. In comparison, the cost of the only retransmission scheme is relatively higher. In other words, this method offersan efficient solution to improve the handover performance. Different FEC schemes have different error correcting capabilities. For the same type of FEC codes, the more the redundant data, the higher the error correction capability. For a given MS velocity, we can estimate the size of FEC codes in order to achieve the desired successful handover probability. And for the sake of simple, we can use the average successful message probability the higher the velocity of an MS, the more the number of bits needed in an adaptive FEC scheme to achieve the target successful handover probability. For example, to achieve a probability of 80%, 2, 10 and 28 bits have to be added into each handover signaling message when the MS is moving at a speed of 50, 70 and 90 km/h, respectively. Then, to achieve the target probability, we can adopt an adaptive FEC scheme that uses a different redundant bit size according to the MS velocity.

2.1.SOFT HANDOVER
The soft handover, in contrast to hard handover, establishes multiple connections with neighboring cells. Soft handover is used by the code division multiple access (CDMA) systems where the cells use same frequency band using different code words. Each MS maintains an active set where BSs are added when the RSS exceeds a given threshold and removed when RSS drops below another threshold value for a given amount of time specified by a timer. When a presence or absence of a BS to the active set is encountered soft handover occurs. The systems using soft handoff are Interim Standard 95 (IS-95) and Wideband CDMA (WCDMA).

III.   METHODS OF SOFT HANDOVERS IN WIMAX
A. **Macro Diversity Handover (MDHO)**The MDHO supported by MS and by BS, the "Diversity Set" is maintained by MS and BS. The Diversity Set is a list of the BSs, which are involved in the handover procedure. The Diversity Set is maintained by the MS and BS and it is updated via MAC (Medium Access Control) management messages. A sending of these messages is usually based on the long-term CINR (Carrier to Noise plus Interface Ratio) of BSs and depends on two thresholds: Add Threshold and Delete Threshold. Threshold values are broadcasted in the DCD (Downlink Channel Descriptor) message. The Diversity Set is defined for each MS in the network. The MS continuously monitors the BSs in the Diversity Set and defines an "Anchor BS". The Anchor BS is one of the BSs from Diversity Set in MDHO. The MS is synchronized and registered to the Anchor BS, further performs ranging and monitors the

downlink channel for control information. The MS communicates (including user traffic) with Anchor BS and Active BSs in the Diversity Set[1]

B. Fast Base Station Switching (FBSS) We are considering fast base station switching technique. In this method a diversity set is maintained for each mobile station. The serving base station and mobile station monitors the neighboring base stations that can be added in diversity set. Diversity set is maintained by both mobile station and serving base station. Diversity set is collection of base stations that can chosen as target base station for a handover. The mobile station selects one base station from diversity set as anchor base station sends its current location to it which is sent to base station controller for decision of a handover. Whenever there is a need of handover base station controller sends handover initiation message to mobile station. Handover decision can be taken by mobile station, base station or base station controller depending upon the implementation [1].

IV. HIERARCHICAL NEIGHBOR DISCOVERY SCHEME

This section describes our proposed scheme and how it supports an optimized MN mobility performance. We argue that a solution considering multiple networks and operators has to contemplate a hierarchical splitting of the existing information. This is due to fact that amount and detail of information pertaining to specific PoAs of single access network, and the combination of all these details for a number of access networks, and details for a number of access networks and different operators, may be very large.The IEEE 802.21 allows the MN to restrict the response message size by optionally setting the MaxResponseSize parameter in the query message. When the response message exceeds the maximum size, some information must be removed from the MIIS response. Clearly, this is not suitable for the user. Removing important information may cause a sub-optimal handover decision. Considering this, and in order to improve the MIIS response in quality, we propose a hierarchical neighbor discovery scheme in which the network hierarchical neighbor discovery scheme in which the network coverage area is divided into mobility zones, managed by different MIIS servers. From bottom to up, the first level of hierarchy is composed by mobility zones defined by the amount of existing networks, users, while even considering areas where networks are overlapped. In the second level, there are Zone MIIS servers (ZMIIS) which are in charge of supplying highly detailed information about specific PoAs in a particular mobility region. The third level refers to the local MIIS servers (LMIIS) managing information of different mobility zones, which belong to the same operator. Finally, a global MIIS server (GMIIS) is specified to be used in multi-operator environments.

Whenever a MN wishes to obtain information regarding the surrounding networks, it sends a MIH Get Information request message to its ZMIIS server. The MN is able to send this message when it detects a new network or when the signal level has crossed pre-defined thresholds. In this work we have proposed for the first method of triggering the MIIS query message. The MN sends a request message to the ZMIIS server through the current PoA link. If the query is related to an entity outside that zone, it is forwarded to the LMIIS server which is able to contact the target zone's ZMIIS and obtain the required information. In the case that the request zone belongs to another operator, the LMIIS server towards the message to the GMIIS server, acting as an interface pointer between relevant mobility regions of different operator. In this way, to which it replies using the MIH get information response message. In case none of MIIS server store information about the detected PoA, the GMIIS server replies with null MIH Get Information response message.Accessing critical information from other operators through non-secure links and third party servers raises important security issues. Other than service agreements, the LMIIS servers must be able to access Authentication, Authorization, and accounting frameworks where users can be authentication prior to do query. One solution in secure inter-domain handover is presented in [2]. We also consider that a node can obtain direct network information without authentication, but in case the information MN receives is minimal.

V. CONCLUSION

The use of FEC with retransmission to protect the handover signaling messages can improve the probability of a successful handover. To achieve the target successful handover probability at different velocities, an adaptive FEC scheme that adaptively adjusts the number of redundant bits according to the MS velocity.Although, using link-going down mechanism will dramatically reduce the handover latency, it is still a challenge to achieve the full mobility: up to 120 km/h, handover latency of less than 50 ms with an associated packet loss that is less than 1 percent. By using the hierarchical view of MIIS server is able to provide the terminal with optimized handover choices.

REFRENCES

[1]   IEEE 802.16 Working Group, "IEEE Standard for Local and metropolitanArea networks, Part 16: Air Interface for Fixed Broadband WirelessAccess Systems," IEEE Std. 802.16-2004, May 2005.

[2]   IEEE 802.16 Working Group, "IEEE Standard for Local and metropolitanArea networks Part 16: Air Interface for Fixed and Mobile BroadbandWireless Access Systems Amendment 2," IEEE Std. 802.16e-2005,Feburary 2006.

[3]   G. P. Pollini, "Trends in handover design," Communications Magazine,IEEE, vol. 34, pp. 82-90, 1996.

[4]   M. G. Williams, "Directions in media independent handover," IEICE Trans. Fundamentals, vol. E88-A, pp. 1772-1776, July 1, 2005.

[5]   D. J. Wright, "Maintaining QoS during handover among multiple wireless access technologies," in Proceedings of Intl. Management ofMobile Business, 2007, pp. 10-10.

[6]   R. Hsieh, Z. G. Zhou, and A. Seneviratne, "S-MIP: a seamless handoff architecture for mobile IP," in Proceedings of INFOCOM , 2003, pp.1774-1784.

[7]   L. Doo Hwan, K. Kyamakya, and J. P. Umondi, "Fast handover algorithmfor IEEE 802.16e broadband wireless access system," in Proceeding ofInternational Symposium on Wireless Pervasive Computing, 2006.

[8]   S. Minsik, K. Hwasung, and L. Sangho, "A fast handover mechanismfor IPv6 based WiBro system," in Proceedings of ICACT,2006.

[9]   Zhiwei Yan, Lei Huang,, C.-C. Jay Kuo, "Seamless High-Velocity Handover Support in
Mobile WiMAX Networks",Xi'an Jiaotong University, Xi'an 710049, P. R. China.

[10]   Zdenek Becvar, Jan Zelenka, Implementation of Handover Delay Timer into WiMAX, (http://fireworks.intranet.gr).

[11]   IEEE P802.16e/D12, "Air Interface for Fixed and Mobile Broadband Wireless Access Systems: Amendment for Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands", October 2005.

[12]   Arpan Mandal, Mobile wimax: pre-handover optimization using hybrid Base Station selection procedure, University of Canterbury, 2008.

[13]   Andrews, J. G., A. Ghosh, Fundamentals of WiMAX : Understanding broadband wireless networking, Prentice Hall ,et al. (2007).

[14]   T. Bchini, N. Tabbane, E. Chaput, S. Tabbane and A-L. Beylot, Performance of Soft Handover – FBSS Compared to Hard Handover in case of High Speed in IEEE 802.16e for Multimedia Traffic, 5th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications March 22-26, TUNISIA 2009.

[15]   IEEE Std: "Air Interface for Fixed and Mobile Broadband Wireless Access Systems," IEEE 802.16e, Part 16, February 2006.

[16]   Rambir Joon, Sandeep, Manveen Singh Chadha,"Analysis of WIMAX Handover",
International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-3, July 2012.

[17]   Fabio Buiati, Luis Javier Garcia Villalba, Daniel Corujo, Joao Soares, Susana Sargento, Rui L. Aguiar," Hierarchical Neighbor Discovery Scheme for Handover Optimization", IEEE communication letters, vol 14, no. 11,November,2010

Active Cancellation Algorithm for Radar Cross Section Reduction

Isam Abdelnabi Osman[1,] Mustafa Osman Ali[2] Abdelrasoul Jabar Alzebaidi [3]

[1, 3] *Electronics Engineering School,* [2] *Electronics & Communication Engineering Department*
[1, 3] *Sudan University of Science and technology,* [2] *Osmania University.*
[1, 3]*Khartoum - Sudan,* [2]*Hyderabad - India*

ABSTRACT:

Abstract - Modern components for signal processing make it possible to achieve radar visibility reduction, that requires reduce the radar cross section (RCS) of an aircraft or a system because it seems to be on the enemy's radar detection capabilities. To achieve this goal, this paper proposed an Active cancellation algorithm for radar cross section reduction using MATLAB, C language program, digital radio-frequency memory (DRFM), and phased array technology to generate the desired signal to cancel the reflected radar returns. The algorithm depends on a pre calculation approach in which an omni direction RCS, clutter, and noise databases generated in advance. Signal processing system function analysis parameter of the measured radar signal. Then find the corresponding echo data (amplitude and phase parameters of the coherent echo) in the target RCS database through real-time amendment. Through the establishment of a target scattering field with the abolition of a coherent signal in the direction of the radar system detection, the radar receiver stays in empty pattern synthesis. The result achieved by the proposed method improves visibility reduction by 25% compared to conventional methods.

KEYWORDS: *Active cancellation, coherent, Echo, radar cross section, MATLAB, Phased array antenna, and Stealth.*

I.   INTRODUCTION

Radar cross section reduction techniques generally fall into one of four categories [1]; target shaping, materials selection and coating, passive cancellation, and active cancellation. The phased-array antenna techniques, high-speed microelectronic devices, and computer processing have made active cancellation technique more feasible and practical. An active cancellation algorithm for radar cross section reduction can readapt to protect any object, such as aircraft. An active cancellation algorithm for radar cross section reduction uses the coherent signal interference. To avoid target detection, the target must transmit a cancellation signal at the same time with an incoming pulse, providing the required phase and amplitude to cancel the reflected energy from detecting radar. The difficulty in implementing such a system is the need to obtain the parameters of cancellation signal in real time, and to achieve precise adjustment of the phase and amplitude of the cancellation waveform. Active cancellation algorithm for radar cross section was based on adaptive real-time adjustment of electromagnetic (EM) signal within a three-dimensional space. The echo signals are produced by the target, when a radar target is illuminated. According to Electromagnetic inverse scattering theory, if the source of radiation field distribution is known, scatter characteristics and distribution of the scattering can be known. If the radar signals are limited within a small precise angle for the EM wave cancellation, the target can be invisible to radar's receiver system. An important part of the development of Active cancellation system for radar cross section understands a particular goal, which is the comparison between the energy density scattered on the radar receiver with incident energy density in the target. The formal definition of the RCS [2],[3],[4] is in (1):

$$\sqrt{\sigma} = \lim_{R \to \infty} 2\sqrt{\pi R [(E_S . \hat{e}_r) / E_i]} \qquad (1)$$

Where $\sqrt{\sigma}$ is the target RCS complex root, E_i is the electric field strength of the incident signal on the target, R is the distance between the target and the radar, $\hat{e}_r$ is aligned unit vector along electric polarization of the receiver, and $\bar{E}_S$ is the vector of the scattered field. Using active cancellation means, reducing the strength incident field on the target to reduce the reflected power to the radar receiver. A target's RCS can be reduced by reducing the target scattering intensity. According to (1); target's RCS can be measured for many scattering directions and a radar target's scattered field direction can be identified as in (2):

$$E_S = \lim_{R \to \infty} [E_i \sqrt{\sigma . \hat{e}_r} \ \big/ \ 2\sqrt{\pi R}] \qquad (2)$$

II.   METHODOLOGY

Active cancellation algorithm for radar cross section reduction consists of two elements:

A.   Hardware components:
- Receiving antenna, used to receive a radar signal.
- Transmitting antenna, used to transmit a cancellation signal.
- Reconnaissance receiver used to stores a precision copy a received signal with aide of (DRFM).

B.   Software(MATLAB/C functions and databases):
- Signal processing and control function (SPC), used to storing a received signal, database searches, signal analysis, processing, and control of other elements.
- Doppler frequency shift modulation function (DFSM), used to superimposes the Doppler frequency on the output signal.
- Power synthesis and beam forming function (PSBF), used to form the modified beam to transmit.
- Target RCS database, is related to frequency, direction, and polarization, power, for incident signal.
- Noise database, is related to the effective noise temperature, input noise power, etc.
- Clutter database, is related to aircraft speed, airborne, carrier frequency, radar point, altitude radar, distance to target and radar pulse repetition frequency (PRF).

The method based on generation an anti phase electromagnetic signal to a target's scattered signal. The effectiveness of this method depends on the knowledge of its real-time characteristics, the measurement precision of the radar signal, and the accuracy of the generated cancellation signal, among other factors. Fig. 1 shows the principle process of the algorithm and Fig. 2, shows the flowchart of an active cancellation algorithm for radar cross section reduction. The incident radar phase, amplitude, frequency, polarization, radar space position and waveform characteristics are accurately and quickly measured on the target platform by SPC, and a reconnaissance receiver. The characteristics of target reflection that correspond to the incident radar waveform will be extracted from the target's RCS database controlled by the computer processing system. By generating a signal (waveform) with the appropriate parameters, including phase, intensity, polarization and frequency, the target's echo can be cancelled when the wave returns to the radar receiving antenna. If we can solve the target to separate N scattering centers, then a radar return on a specific frequency is (3):

$$\sigma = \left| \sum_{n=1}^{N} (\sigma_n)^{0.5} . e^{j\varphi_n} \right|^2 \qquad (3)$$

Where σ_n is the Nth scatter RCS and φ_n is the phase due to the physical location of scatterer's.
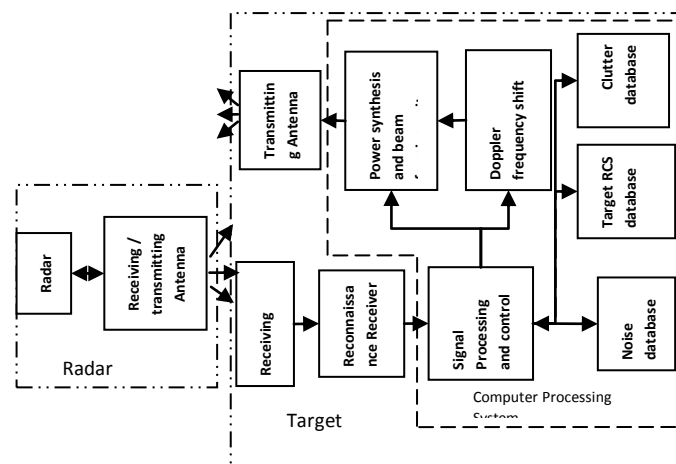
Fig. 1: The principle of an active cancellation algorithm for radar cross section reduction process.
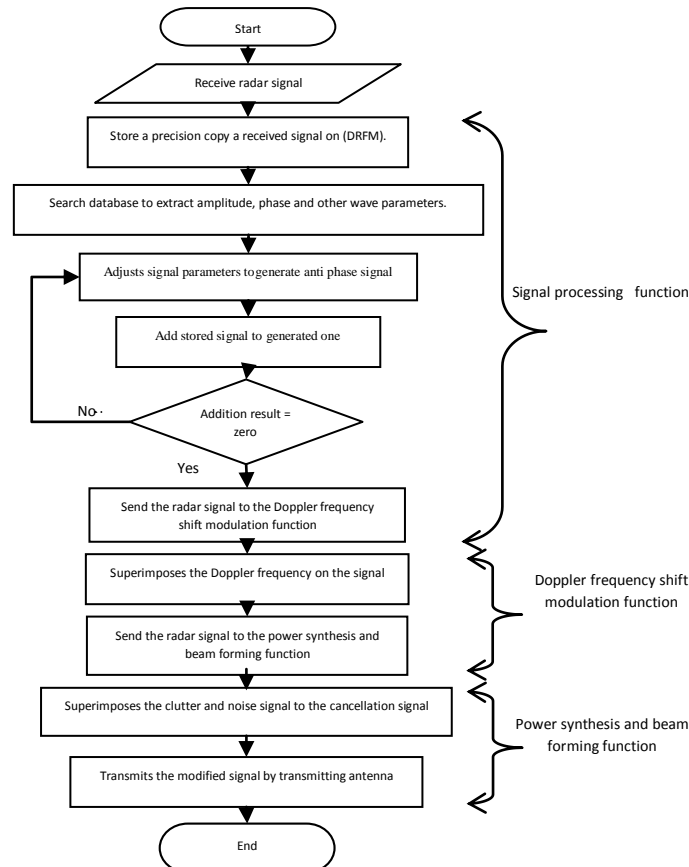
Fig. 2: An active cancellation algorithm for radar cross section reduction flowchart.

A target with a huge number of scattering centers, several controlling scattering centers will exist for a specific incident signal angle and operating radar frequency. Reduction of the radar returns from these centers can reduce the target's RCS. If the original RCS of the target is denoted as σ_0, an active cancellation system for RCS reduction can produced an equivalent RCS scattering center denoted by σ_1. The phases are φ_0 and φ_1, respectively of scattering centers. The superposition $\sigma 0$ and $\sigma 1$ is given by (4):

$$\sigma = \left| (\sigma_0)^{0.5} e^{j\varphi_0} + (\sigma_1)^{0.5} e^{j\varphi_1} \right| \qquad (4)$$

By analyzing (4), (5) will get as following:

$$\sigma = \sigma_0 \left| 1 + \sigma_1 / \sigma_0 + 2(\sigma_1/\sigma)^{0.5} [\cos(\varphi_1 - \varphi_0)] \right| \qquad (5)$$

Controlling σ_1 and φ_1 can optimize those parameters to get (6):

$$\begin{cases} \sigma_1 = \sigma_0 \\ \varphi_1 - \varphi_0 = (2k+1) \quad (where \quad k \ is \ Integer) \end{cases} \qquad (6)$$

When $\sigma = 0$, this indicates that stealth in the direction of the enemy's radar has been achieved.

Compiling the target's RCS database is an important step in designing an active cancellation algorithm for RCS reduction. The RCS entry is a function, rather than just varied number with different frequency, direction, and polarization for incident signal. It is necessary to establish an RCS database due to different frequencies, polarizations, and directions, according to real-time measurements for frequency, direction, power, and polarization of incident signal. This database must support real-time modification for the parameter of the transmitter to produce an effective cancellation signal for transmission.The reconnaissance receiver is used for reception and reconnaissance signals from enemy's radar transmitters. A received radar signal is applied to SPC function which stores a precision copy a received signal with aide of (DRFM) and alternately, it analyzed for amplitude and phase adjustment to generate anti phase signal with same amplitude. By comparing the stored signal with the wave signal generated by SPC function, the result from this comparison is checked to obtain a minimum output value (zero balance), when this happen, the SPC function will send the radar signal to the DFSM function, with coherent superposition of clutter and noise. The transmitting antenna and PSBF function are used to form and transmit the active cancellation signal.

The system's memory is used for storing the databases, including the noise database, target echo database, and the clutter database. The algorithm work with the assumption that an echo signal consists of three parts: noise, clutter, and target echo so a radar echo can be as follows (7):

$$X(t) = S(t) + N(t) + C(t) \qquad (7)$$

Where S(t) is target's echo signal, N(t) is the noise signal, and C(t) is the clutter signal.

Due to a great deal of calculation and processing power that required to determine the radar's cancellation signal; it is difficult to achieve calculations in real-time without pipeline delays. For this, an offline calculation is used to build a target RCS database. Approximation prediction method is the based method for obtaining a complex target RCS [5],[6],[7].The database for noise and clutter usually uses a distribution of Gaussian for White noise, which can be produced by the Monte Carlo method [8][9]. Clutter data is related to aircraft speed, airborne, carrier frequency, radar point, altitude radar, distance to target and radar pulse repetition frequency (PRF). When clutter data are calculated, reduction of large amount of data is done because the speed, radar frequency, and aircraft altitude are fixed, and only PRF and radar point are changed [10].Signal processing and control function is the core of an active cancellation algorithm for radar cross section reduction, which used for storing a received signal, database searches, signal analysis, processing, and controlling the other elements. The Doppler shift in the radar wave is produced due to superimpose of the Doppler frequency by the Doppler frequency shift modulation function. The power synthesis and beam forming function, used to form the modified beam to transmitted via the transmitting antenna under control of signal processing and control function.

III.   TEST AND RESULTS:

An active cancellation algorithm for radar cross section reduction was tested with the following conditions for evaluation purpose:
1. A coherent pulse train with 1 MHz modulation rate for radar transmit signal.
2. The signal has PRF of 1 kHz and a pulse width of 4 μs.
3. Uniform speed for the target movement with 100 km initial distance away from the radar transmitter, 300 m/s initial radial velocity, 0 deg of both azimuth and elevation angles and 2 m^2 target's RCS according to the Swelling model II.
4. The reference pattern function for reconnaissance is described by (8):

$$F(\theta) = \begin{cases} ASa & [a\theta / (\theta_1 / 2)]\, k_0 & |\theta| \le a_1 \\ BSa & [a(\theta \pm a_2) / (\theta_2 / 2)]\, k_0 & |\theta| > a_1 \end{cases} \qquad (8)$$

Where $\kappa_0 = (\cos\theta 0)^{0.5}$ is the control factor for phased-array antenna modified beam gain with variation of scanning angle, θ_0 is the beam of scanning angle, θ_1 is the unbiased-beam main-lobe beam width of 3 dB, θ_2 is the unbiased-beam first side-lobe 3-dB beam width, AS is the unbiased-beam main-lobe gain value, BS is the unbiased-beam first side-lobe gain value, a = 2.783, ($a_1 = \pi\theta_1/a$) is first zero unbiased-beam (in rads), and ($a_2 = \pi(\theta_1 + \theta_2/a$) is the unbiased-beam first side-lobe peak point of view (in rads).The 3D Electromagnetic pattern can be simplified into an elevation and azimuth multiplication pattern result as shown in (9).

$$F(\theta, \varphi) = F_\theta(\theta) F_\varphi(\varphi) \qquad (9)$$

Where $F_\varphi(\varphi)$ is the elevation pattern and $F_\theta(\theta)$ is the azimuth pattern. Assuming that the radar antenna vertical main-lobe beam-width is 2 degree, the main-lobe gain is 40 dB, the gain is 9 dB, and the first side-lobe-width is 1 degree, the 3D antenna pattern shown in Fig. 3 will be generated.
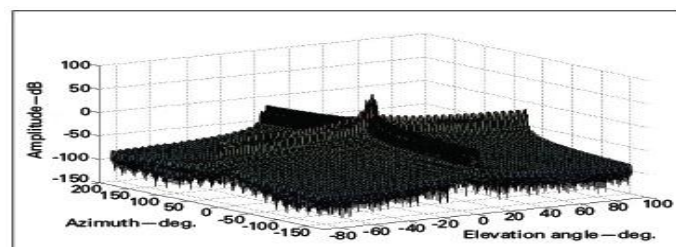
Fig. 3: Three dimensional (3D) antenna's pattern.

The algorithm can used for (ECM) Electronic countermeasures function using rectangular array antenna M x N pattern function described below as (10):

$$g(\theta, \varphi) = G(\theta, \varphi) \left| E(\theta, \varphi) \right| \left| e(\theta, \varphi) \right| \qquad (10)$$

Where g(θ, φ) is the pattern of the antenna, G(θ, φ) is the factor of the directivity, E(θ, φ) is the array factor for the beam shape determines, e(θ, φ) is the factor of the array element with (e(θ, φ) ≈ 1), φ is the elevation angle

on spherical coordinates array, θ is the azimuth angle on spherical coordinates array, (φ ∈ [0, π/2]), and(θ ∈ [0, 2π]). Adjacent-array element spacing of d = λ/2 can be described in the x and y directions, E(θ, φ) as (11):

$$E(\theta, \varphi) = \sum_{m=1}^{M} \sum_{n=1}^{N} I_{mn} \exp[\, jkd \,(m \tau_x \, n \tau_y)] \qquad (11)$$

Where k = 2π/λ is the wave number and I_{mn} is the weighting coefficient.

$$\begin{cases} \tau_x = \sin\theta \cos\varphi - \sin\theta_0 \cos\varphi_0 \\ \tau_Y = \sin\theta \cos\varphi - \sin\theta_0 \cos\varphi_0 \end{cases} \qquad (12)$$

Where (θ_0, φ_0) is a beam pointing vector. If M = 51, N = 21, θ_0 = 30 deg., φ_0 = 20 deg, the (51 x 21) will result array antenna pattern shown in Fig. 4.
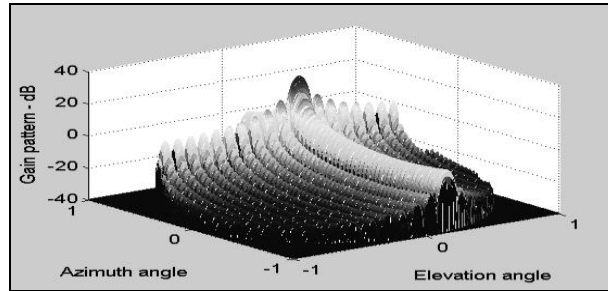
Fig. 4: Pattern from multiple-element array antenna.

Fig. 5, shows the result obtained from the algorithm, (a) shows the a coherent pulse train spectrum, (b) shows superimposed of coherent pulse train on the noise and clutter waveform, with completely target signal submerged under noise and clutter, (c) shows a signal spectrum contains added noise and clutter with the target signal, and (d) shows the contrast before (top) the cancellation signal and (bottom) after the cancellation signal has been added to radar return.

The cancellation signal can be described as (13):

$$S = 20 \log[\, 1 + \left| \Delta \vec{E} / \vec{E}_S \right| \,] \qquad (13)$$

Where $\Delta \bar{E}$ is the cancellation residual field and $\bar{E}_S$ is the target's scattering field. Complete stealth is realized when S = 0, From Fig. 5 (d), it can be seen that corresponding cancellation signal, S to $\Delta \bar{E}$ max (6 x 10−2dB) is 0.51 dB, so the reduction of radar detection maximum range is 25% of original value.

IV.   CONCLUSION:

From the results, an active cancellation algorithm for radar cross section reduction reduces the target detectability. This approach can be used with different number of others radio echo scenarios.
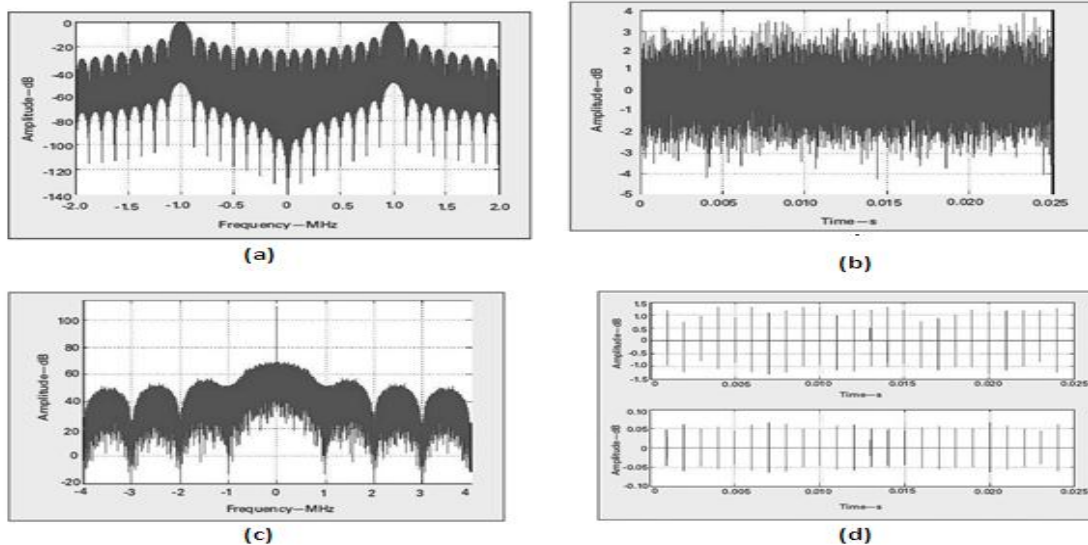
REFERENCES

[1]   Eugene F. Knott, John F. Shaeffer, and Michael T. Tuley, Radar Cross Section, SciTech Publishing, Inc., 2004.
[2]   David C. Jenn, Radar and Laser Cross Section Engineering, American Institute of Aeronautics and Astronautics, Inc.2005
[3]   Bassem R. Mahafza, Radar Systems Analysis and Design using MATLAB, Chapman& Hall/CRC, 2000.
[4]   Merrill I. Skolnik, Radar Handbook, McGraw-Hill, third edition, 2008.
[5]   Michael O. Kolawole, Radar Systems, Peak Detection and Tracking, Newns, 2002.
[6]   Fred E. Nathanson and J. Patrick Reilly, Radar Design Principles, McGraw-Hill, 1999.
[7]   V. G. Nebabin, Methods and Techniques of Radar Recognition, Artech House, 1995.
[8]   David K. Barton and Sergey A. Leonov, Radar Technology Encyclopedia, Artech House, 1998.
[9]   Yakov D. Shirman, Computer Simulation of Aerial Target Radar Scattering, Recognition, Detection, and Tracking, Artech House, 2002.
[10]   Merrill I. Skolnik, Introduction to Radar Systems, McGmw-Hill, 2001.

Absorption of Light in Silicon Nanowire Solar Cells: Designing Of Solar Cells

Prof. R.L. Sharma[1], Juhi Narain[2]

Dept. of Electronics & Communication Engineering,
Ajay Kumar Garg Engineering College

ABSTRACT

In the proposed study, an investigation has been carried out in order to find a material efficient structure, capable of harnessing maximum solar spectrum. A material efficient structure designed using a thin film and Si nanowire. Silicon material is used as it leads to environmental friendly design. The principal objective of this study is to maximize the photon absorption, keeping reflection to a minimum. Such reflection are caused due to the regular surface of the thin film. As a thumb rule, the performance of the solar cell can be elevated by increasing the photon absorption phenomenon in the Si material. The proposed approach may yield an efficient hybrid structures (involving the combination of nanowire and thin film). Using such geometry, it is expected to achieve a higher efficiency (increased EHP generation), compared to nanowires or thin film alone. It is expected to achieve a higher Electron Hole Pair (EHP) generation or higher performance efficiency because of the two following reasons. First, Si nanowire exhibits low reflection due to low reflection area and the other being that such hybrid structure offers more active surface area for the EHP generation. Using this structure an internal quantum efficiency (IQE) upto 21% can be achieved which is quite good for commercially used solar cells.

***KEY WORDS**: Solar spectrum, Silicon nanowire, Electron Hole Pair (EHP) Generation, Internal Quantum Efficiency (IQE).*

Section I

I.   INTRODUCTION: NEED OF SOLAR ENERGY

Energy has been an integral part of one's life. It is a source of living for all creatures. Energy has found various applications in transportation, industries, agriculture, households and offices. In the pre-industrial area, energy was most commonly available to human in the form of fossil fuels. In 1970s, due to the limited availability and increase in cost of fossil fuels, need to find other forms of energy occurred (6,7,8). Although energy can have many forms like heat energy, electrical energy, nuclear energy and so on but sun's energy due to its abundance

and sustainability was the most favourite among the researchers. Apart from sunlight and solar heating, the energy of the sun is also available in the form of biomass and wind energy thus providing security for the future development and growth.

Section II

II.   DESIGNING OF SOLAR CELL

Solar cells have been one of the best device in storing and utilizing the sun's light. In this paper, we have tried to design a solar cell using silicon material only. Only silicon was used because our aim was to make the structure cost efficient and Si on earth, after oxygen is the next material found in abundance. Various researchers have used aluminium doped zinc oxide (ZnO:Al) for fabrication purpose (3). Our designed structure may yield an efficiency upto 21% which is quite high compared to other designed cells using transparent conductive oxide, TCOs**.**   In initial designing of a solar cell we had two options either to design a cell cuboidal in shape using Si material or to use a Si nanowire (NW) over this cuboidal structure. In both the designs only silicon material is used. The volume $(.205\mu m^3)$ was kept constant in both the designs (4). Also the solar spectrum of wavelength 400nm to 1100nm (AM1.5) was chosen. At this air mass (AM) the sun is exactly above

the earth so chances of maximum absorption is increased. The thickness of the cell is not exceeded above 2500nm. The thickness has to be kept as less as possible for maximum absorption of photons.

Our main aim was to design a structure which is cost efficient as well as provide increased quantum efficiency (QE). The quantum efficiency is defined as the probability of an incident photon that results in generation and collection of an electron in the solar cell (1,2). There are two types of Quantum Efficiency are (a) External Quantum Efficiency (EQE). (b) Internal Quantum Efficiency (IQE). The photons that are absorbed results in generation and collection of electrons, which contribute to the solar cell current, thus IQE is used to study solar cell performance.

$$QE(\lambda) = \frac{Pabs\ (\lambda)}{Pin\ (\lambda)} \qquad\qquad Eq\ 1$$

$P_{abs}(\lambda)$ is power of the absorbed light and $P_{in}(\lambda)$ is power of the incident light. Internal Quantum Efficiency (IQE) is given as

$$IQE = \frac{Number\ of\ photons\ absorbed\ by\ the\ solar\ cell}{Number\ of\ photons\ falling\ onto\ the\ solar\ cell} \qquad\qquad Eq\ 2$$

Section III

III. SOLAR CELL USING ONLY SI THIN FILM

In designing of the solar cell using only Si thin film, a cuboid of dimensions (2500nm*2500nm*320nm) was taken. An illuminated source similar to that of solar spectrum within the range 400nm to 1100nm was applied over this thin film. As solar light consists of photons, these photons when fall upon the film get absorbed, reflected and transmitted (5). The %power photons transmitted, absorbed and reflected with respect to the wavelengths is shown in the figure 1(a), 1(b) and 1(c) respectively. For measuring the above values various DFT monitors were used on several places in the designed solar cells, especially above the film to measure the reflection, at the bottom of the film to measure the transmission and in between to determine the absorbed photons. As photons are a source of energy, the graph is determined between %power and wavelength. The overall energy is assumed to be 1. Thus once the transmission and absorption is determined, reflection can be found using R=1-A-T (10,11).In designing this structure the main concern was whether the structure is correctly designed or not, and whether the photon absorption will be maximum. For this the absorption graph obtained was compared with the absorption curve obtained when Si was fabricated with ZnO (3). On comparison it was found that the curve obtained was nearly similar within the spectrum range 0.6 to 0.9 µm.
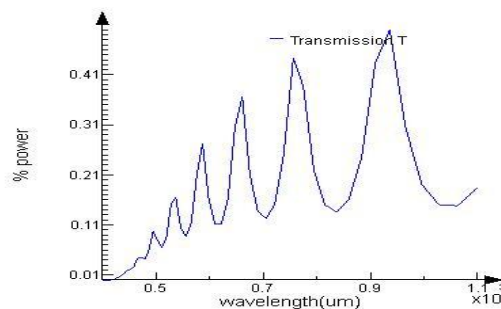
Fig.1(a). Percent power transmitted wrt wavelength range 400nm to 1100nm

The variations in other wavelengths is due to the TCO used in the reference curve. This showed us that the designed cell using Si thin film was nearly perfect to the fabricated design using ZnO. The internal quantum efficiency which depends upon the photons absorbed came out to be 17%-19%.
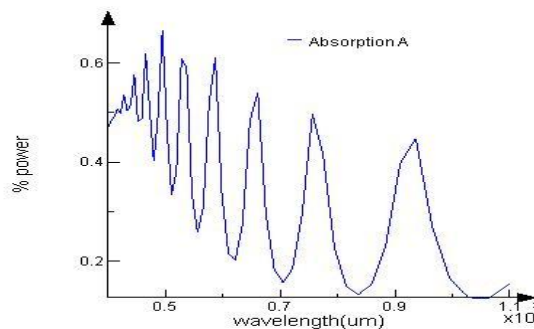
Fig1(b). Absorption curve for photons entering the solar cell
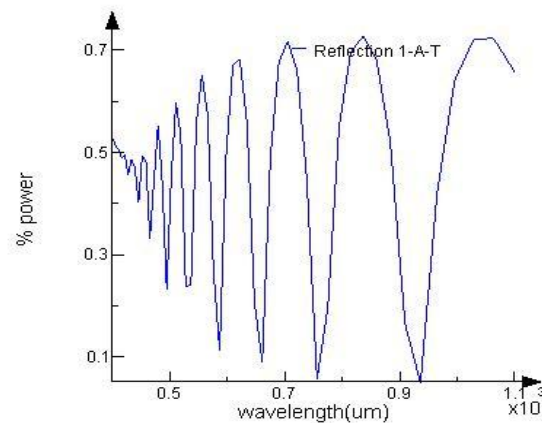
Fig1(c). Reflection of photons within the solar spectrum

3.1.Solar Cell designed using Si thin film and Si nanowire(NW)

A more efficient solar cell can be designed only by increasing the optical length of the solar cell. Optical length is the total length which the photon travels inside the solar cell. As volume and the chosen material is our most important constraint we designed a structure in which a Si nanowire of radius 35nm and height 1500 nm is taken. The dimensions of the Si thin film is reduced accordingly. The main advantage in this design is that as the thickness of the thin film is reduced, probability of photon absorption is increased, also a longer optical path is achieved. Using the above designed solar cell, an internal quantum efficiency of 21% can be achieved, which is higher than the structure designed using only Si nanowire. The absorption curve for the designed structure is shown in Fig2.
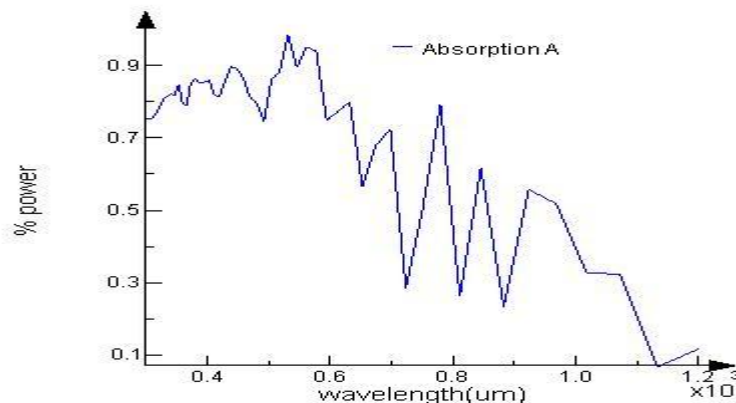
Fig2. Absorption curve for Si thin film and nanowire designed structure

Section IV

IV. CONCLUSION

Out of the two designed studied above we found that increase in the optical length of the solar cell plays an important role in increasing the photon absorption. Thus, the structure with Si film and nanowire came out to be most efficient as its optical length was increased. The designs are cost efficient as only Si has been used which only requires purification. Keeping volume constant and reducing the thickness of the substrate provided us a structure of Si thin film with Si nanowire which gave us the IQE of 21%. The comparison of the two designed structure showed us the absorption curve of photons within the solar cell. Through this absorption curve IQE was determined and hence the better structure was achieved.

ACKNOWLEDGMENT

REFERENCES

[1]   Albert Y.C.Yu, 'The metal-semiconductor contact: an old device with a new future', presented in IEEE spectrum March 1970.

[2]   Prof. E.H. Rhoderick, 'Metal semiconductor contacts', presented in IEEE PROC., Vol.129, Pt.I, No 1, February 1982.

[3]   Joachim Muller, Bernd Rech, Jiri Springer amd Milan Vanecek, 'TCO and light trapping in Silicon thin film solar cells', in Solar Energy, www.elsevier.com/locate/solener, Solar Energy77 (2004) Pg 917-930.

[4]   L. Tsakalakos, J.Balch, J.Fronheiser, B.D. Korevaar, O. Sulime et al., 'Silicon Nanowire Solar Cells', Applied Physics Letters,91,233117(2007); doi: 10.1063/1.2821113.

[5]   Erik Garnett and Peidong Yang, 'Light Trapping in Silicon Nanowire Solar Cells', published in 'Nano Letters, American Chemical Society', 2010, Pg.1082-1087.

[6]   A. Shah, P. Torres, R. tscharner, N. Wyrsch and H.Keppner, Science 285, 692 (1999).

[7]   W.U. Huynh, J.J. Dittmer and A.P. Alivisatos, Science 295, 2425 (2002).

[8]   B.M. Kayes, N.S. Lewis and H.A. Atwater, J.Appl. Phys. 97, 114032 (2005).

[9]   H.B. Nie, S.Y. Xu, S.J. Wang, L.P. You, et al., Appl. Phys. A: Mater.Sci. Process 73, 229 (2001).

[10]   C. Beneking, B. Rech, S. Wieder, O. Kluth, et al., 1999, 'Recent Development of Silicon thin film solar cells on a glass substrate.' Thin Solid Films 351, Pg. 241-246.

[11]   X. Deng, K.L. Narsimhan, 1994, 'New evaluation technique for thin film solar cell back reflector using photothermal deflection spectroscopy', in proceedings of the Ist World Conference on Photovoltaiv Solar Energy Conversion (WCPEC), Hawaii, pp(555-558).

A Novel Approach for High Performance Slew Rate

Lalit Mishra, M Tech 4th Sem,
Nitin Meena ,Assistant Professor

Department of VlSI And Embedded System IES College of Technology, RGTU ,Bhopal

ABSTRACT

In this paper a research is proposed to enhance the slew rate using current mirror circuit and cascaded folded amplifier. Most slew rate enhancement circuits can either be used in current-mirror amplifier or folded-cascade amplifier, but not in both amplifiers. This circuit is implemented on AMS .65µm cmos process using a current mirror circuit with cascaded folded amplifier has very improved slew rate.

***KEYWORD**- Amplifier, Load Capacitance, Slew rate Enhancement circuit, Transient Response*

I.   INTRODUCTION

For the applications of low-power high-speed switched capacitor circuits, fast settling time of an operational amplifier is a common and critical requirement. The settling time of an amplifier can be divided into the slewing period and the quasi-linear period. In particular, the quasi-linear period depends on the small-signal behavior of the amplifier while the slewing period depends on the large-signal behavior. The settling time of these amplifiers is dominated and restricted by its slewing period as the maximum available current Imax to charge up the loading capacitor is limited in low power condition. The slew rate (SR) of single-stage amplifiers is given by

SR = Imax/CL

There are many works proposed to improve the slew rate using the idea of dynamic bias. Degrauwe proposed adaptive biasing based on circuit subtractors and current mirrors with gained ratio on the differential amplifier, so that the adaptive bias circuit is enabled to increase the bias current of the tail current source when there is a transient signal at the input. However, perfect current subtraction cannot be achieved due to mismatch of the current mirror at different operation regions. As a result extra offset voltage is a critical point of this design.[1]

1.1.Single Stage Amplifier- MOS transistors are capable of providing useful amplification in three different configurations. In the common-source configuration, the signal is applied to the base or gate of the transistor and the amplified output is taken from the drain. In the common-drain configuration, the signal is applied to the base or gate and the output signal is taken from the source. This configuration is often referred to as the source follower. In the common-base or common-gate configuration, the signal is applied to the emitter or the source, and the output signal is taken from the collector or the drain. Each of these configurations provides a unique combination of input resistance, output resistance, voltage gain, and current gain.

1.2.Common-Source Configuration

The resistively loaded common-source (CS) amplifier configuration is shown in Fig. 1(a) using an n-channel MOS transistor. The corresponding small-signal equivalent circuit is shown in Fig.1(b). As in the case of the bipolar transistor, the MOS transistor is cutoff for Vi = 0 and thus Id = 0 and Vo = VDD. As Vi is increased beyond the threshold voltage Vt, nonzero drain current flows and the transistor operates in the active region(which is often called as saturation for MOS transistors) when Vo > VGS-Vt.
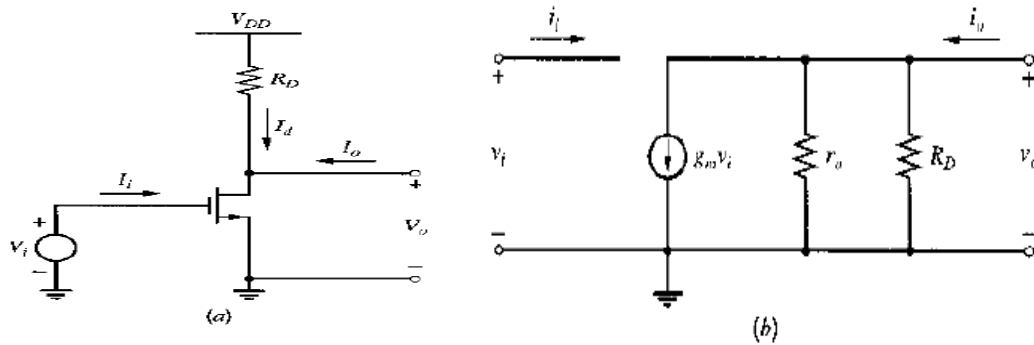
Figure.1 (a) Resistively loaded, common-source amplifier (b)Small-signal equivalent circuit

$$V_a = V_{DD} - I_d R_D$$

$$= V_{DD} - \frac{\mu_n C_{ox}}{2} \frac{W}{L} R_D (V_i - V_t)^2$$

The output voltage is equal to the drain-source voltage and decreases as the input increases. When Vo < VGS – Vt, the transistor enters the triode region, where its output resistance becomes low and the small-signal voltage gain drops dramatically.

Fig.2. shows the voltage characteristics of the circuit.The slope of this transfer characteristic at any operating point is the small-signal voltage gain at that point. The MOS transistor has much lower voltage gain in the active region than does the bipolar transistor, therefore the active region for the MOS CS amplifier extends over a much larger range of Vi than in the bipolar common-emitter amplifier.
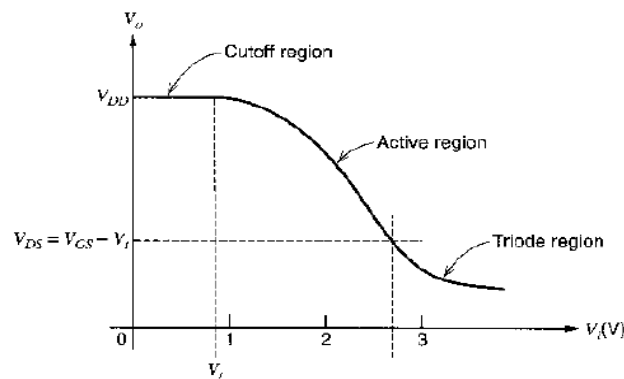
Figure.2 Output voltage versus input voltage for the common-source circuit.

The transconductance G_m is

$$G_m = \left. \frac{i_o}{v_i} \right|_{v_b - 0} = g_m$$

$$R_i = \frac{v_i}{i_i} \longrightarrow \infty$$

Input resistance

Output resistance seen looking into output with input shorted,

$$R_o = \left. \frac{v_o}{i_o} \right|_{v_i = 0} = R_D \parallel r_o$$

The open-circuit voltage gain is,

$$a_v = \left. \frac{v_o}{v_i} \right|_{i_o - 0} = -g_m(r_o \parallel R_D)$$

1.3. Current-Mirror with Static-Bias and Dynamic-Bias

By using the static-bias and dynamic-bias a very high slew-rate current- mirror CMOS op-amp can be designed. It uses a circuit to inject an extra bias current into a conventional source coupled CMOS differential input stage in the presence of large differential signals. This measure substantially increases the slew-rate of an operational-amplifier for a given quiescent current.
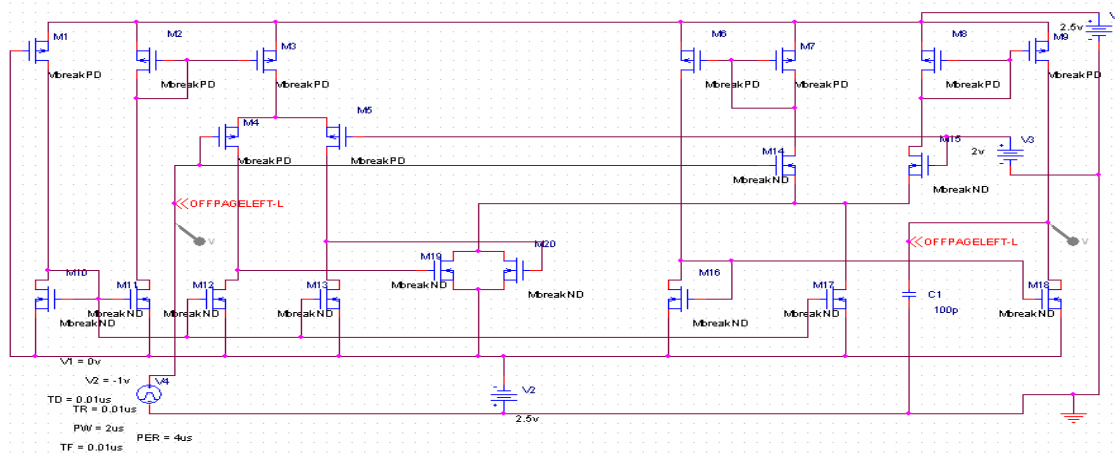
Figure.3. Current-mirror with static-bias and dynamic-bias.

Transient Analysis of current-mirror with static and dynamic bias
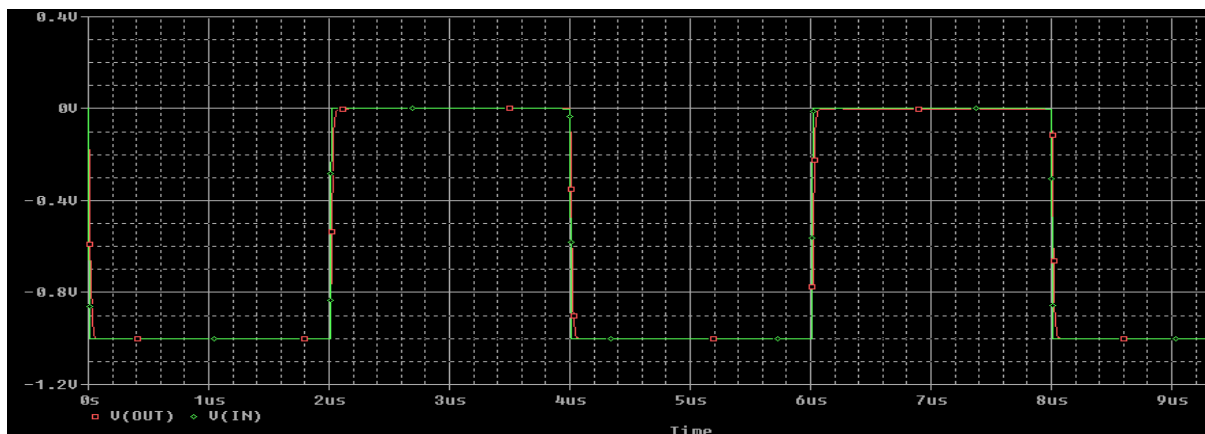
Figure 4 Transient analysis of CM with static and dynamic.

II.   CURRENT-MIRROR SRE CIRCUIT

The technique at the active load device of the core amplifier is used to sense the fast signal transient. The simple SRE circuit is used as a plug-in feature to the core amplifier and do not affect its original small signal frequency response. Fig5,shows the circuit-
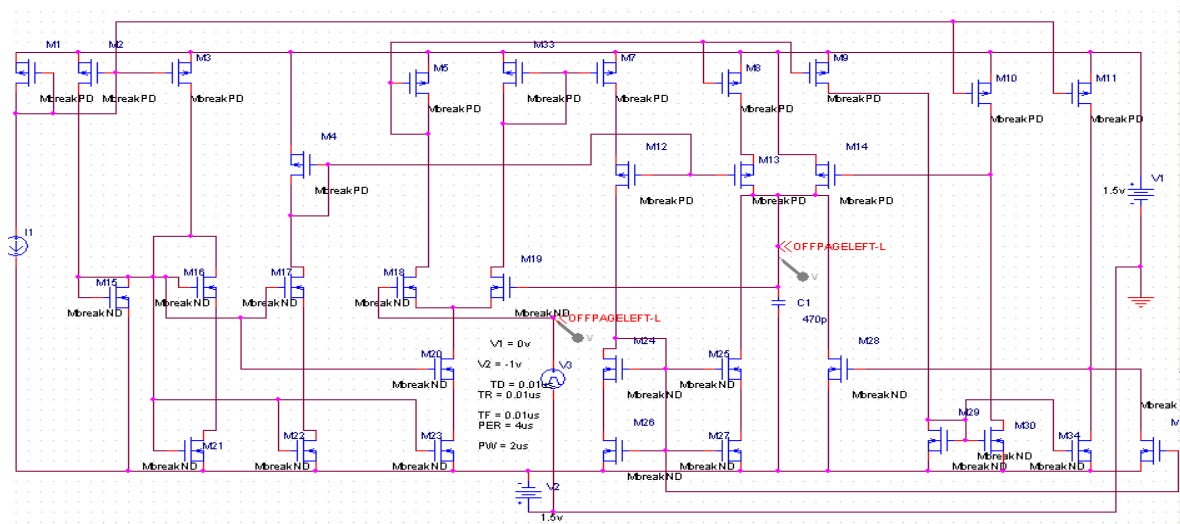
Figure.5 Current-mirror with SRE circuit.

III. CONCUSION

In the future technology has change rapidly, primarily to the larger unity-gain frequency and slew rate, the current-mirror OpAmp may be preferred over the folded-cascode OpAmp. However, one has to be careful that the current-mirror OpAmp has larger input noise as well, as its input stage is biased at a lower portion of the total bias current and therefore a relatively smaller gm given the same power consumption.In the previous work we know that large capacitive load decreases slewing rate if we increase of biasing current which increases the static current loss in the amplifier circuit.

REFERENCE

[1]   M.G.Degrauwe,J.Rijments,E.A.Vittozand H.J.Deman,"Adaptive biasing CMOS amplifiers,"IEEE journal ofsolid state circuits,vol.17,No.3,pp.522-528,jun 1982.

[2].   K.Nagaraj, "CMOS amplifiers incorporating a novel slew rate enhancement technique" in proc.IEEE Custom Integrated CircuitsConference, pp.11.6.1-11.6.5, 1990.

[3].   J.Ramfrez-Angulo,"A novel slew-rate enhancement technique for one-stage operational amplifier,"in proc.IEEEMid west Symposium on Circuits andSystems.pp.11-13, Aug.1996.

[4].   R.Klinke,B.J.Hostika and H.J.Pfleiderer, "A very-high-slew-rateCMOS operational amplifier,"IEEEJournal of Solid-Statecircuits.Vol.24,No.3,pp.744-746,jun.1999.

[4].   H.Lee and P.K.T.Mok,"A CMOScurrent-mirror amplifier with compactslew rate enhancement circuit forlarge capacitive load applications" inproc. IEEE International Symposiumon Circuits and Systems.Vol.1,pp.220-223, 2001.

[5].   J.Ramirez-Angulo,R.G.Carvajal,J.A.Galan and A Lopez-Martin,"A free but efficient lowvoltageclass-AB two stageoperational amplifier," IEEETransactions on Cicuits and SystemsII,Vol.53.No.7,pp.568-571,jul.2006.

[6]   P. R. Gray, P. H. Hurst, S. H. Lewis and R. G. Meyer,Analysis and Design of Analog Integrated Circuits, 4tl' ed.New York: Wiley, 2001

[7]   H. Lee, P. K. T. Mok and K. N. Leung, "Design of Low-Power Analog Drivers Based on Slew-Rate Enhancement Circuits for CMOS Low Dropout Regulators," IEEETransactions on Circuits and Systems-II:Express Briefs, Vol.52, No. 9, pp. 563-567, Sep. 2005

[8]   V. W. Lee and B. 1. Sheu, "A high slew-rate CMOSamplifier for analogsignal processing," IEEE Journal ofSolid-State Circuits, vol. 25, no. 3,pp. 885-889, Jun. 1990.

[9]   Ka Vee Kwong & Ka Nang Leung Slew-Rate Enhancement Circuit of CMOS Current-Mirror Amplifier by Edge-Detecting Technique. 2010 IEEE International Conference of Electron Devices and Solid-State Circuits (EDSSC)

[10]   Ron Hogervorst, RemcoJ.Wiegerink, Peter A.L de jong,JeroenFonderie, Roelof F. Wassenaar, Johan H.Huijsing,*CMOS low voltage Operational amplifiers with constant gm rail to rail input stage,* IEEE proc. pp. 2876-2879, ISCAS1992.

[11]   Giuseppe Ferri and Willy Sansen*A Rail-to-Rail Constant-gmLow-Voltage CMOS Operational TransconductanceAmplifier,* IEEE journal of solid-state circuits, vol.32, pp 1563-1567, October 1997.

[12]   Sander l. J. Gierkink, peter j. Holzmann, remco j. Wiegerinkand Roelof f. Wassenaar, *Some Design Aspects of a Two-Stage Rail-to-Rail CMOS Op Amp.*

[13]   Ron Hogervorst, John P. Tero, Ruud G. H. Eschauzier, and Johan H. Huijsing, *ACompact Power-Efficient 3 V CMOSRail-to-Rail Input/Output Operational amplifier for VLSICell Libraries,* IEEE journal of solid state circuits, vol.29, pp1505-1513,December 1994

[14]   http://www.d.umn.edu/~htang/ECE5211_doc_files/ECE5211_files/Chapter6.pdf

Outage Analysis of Cooperation Over Wireless Network

Sohrab Alam[1], Sindhu Hak Gupta[2]

[1,]*Department of Electronics & Communication Engineering,*
[2,]*A.S.E.T, Amity School of Engineering & Technology, Amity University, Noida, Uttar Pradesh, India.*

ABSTRACT

In this paper, we analyze the performance of cooperative communication system for independent and flat Nakagami-m fading channel parameters by using amplify and forward cooperative protocol. We have estimated the outage probability, channel capacity and ergodic capacity of the system. We further simulate the equations of these parameters by plotting them for verification.

INDEX TERMS: Diversity technique; cooperative communication; cooperative communication protocols; Nakagami-m fading; ergodic capacity; channel capacity; outage probability.

I. INTRODUCTION

An efficient technique of cooperation among the users having single antenna is required way to achieve the transmit diversity in such a system which is unable to support multiple antennas either at the transmitter/receiver side or on both sides due to size, cost, power, complexity of hardware [1], [2]. So, the virtual array of antennas is formed by cooperation in which the source broadcasts the multiple copies of the signal toward the relays and destination with same power. Relays also forward the received copies of the same signal toward the destination as shown in fig. 1. At the destination, we observe the resultant information after maximum ratio combining (MRC) of all the multiple received copies of signal [3], [4], [5].
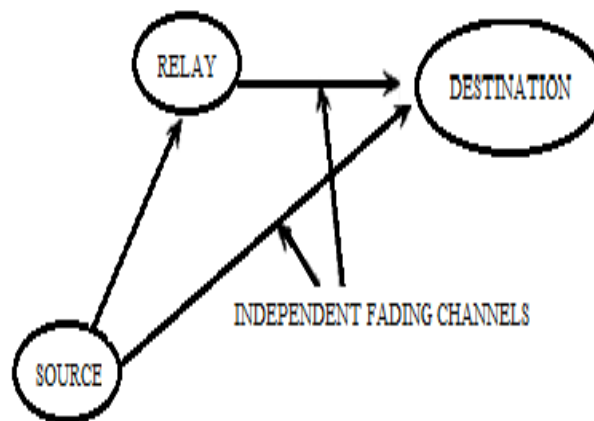
Fig.1   Cooperative communication

There are different types of cooperative communication protocols [6], [7], [8]-:
1. Amplify and forward,
2. Detect and forward, and
3. Coded cooperation.

In this paper, we use amplify and forward cooperative protocol strategy in which relays amplify the received signals and retransmit them toward the destination as shown in figure 2.
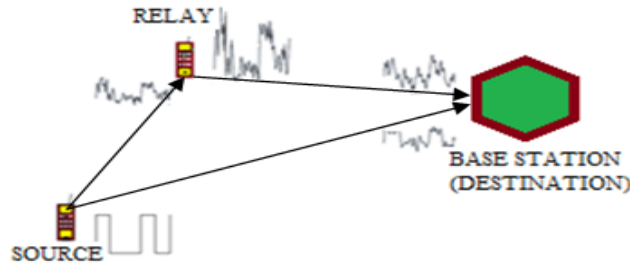
Fig. 2. Amplify and Forward

Here, we have considered Nakagami-m fading channels to be independent and flat with parameter m.

II. SYSTEM MODEL

We have considered a system, which has a source (S), multiple relays i.e. R_1, R_2,R_3, …Rn and the destination (D) as shown in fig. 3.
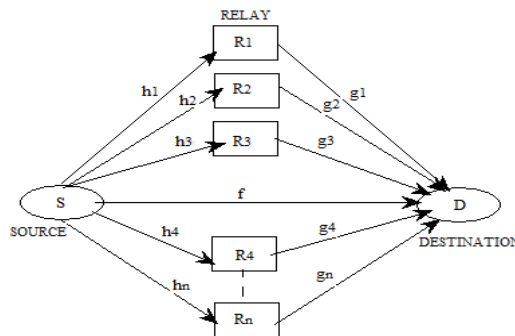
Fig.3. Cooperative communication network

The channel coefficient between source and destination is flat Nakagami-m fading coefficient denoted by f. The channel coefficient between source and i[th] relay is hi and between i[th] relay and destination is gi. These are the flat and independent Nakagami-m fading channels.

Here, i= 1, 2, 3, 4,……,n. The signals are transmitted in two phases.
In first phase, source transmits the signal's one copy directly to the destination and multiple copies of the same signal toward the relays with the same power (i.e. Ps). In second phase, all the relays Ri, where i= 1, 2, 3, 4,……,n, transmit the amplified versions of received signal copies toward the destination with power PRi. Where i= 1, 2, 3, 4,……,n for n number of relays. Hence the total transmitted power toward the destination is as

$$P_T = Ps + \sum_{i=1}^{n} P_{R_i} \tag{1}$$

We also know that the ratio of transmitted power from source to destination (P_S) to the total transmitted power toward the destination (P_T) is called the power distribution ratio i.e.

$$\alpha = \frac{P_S}{P_T} \tag{2}$$

α lies between 0 and 1. The power transmitted by each relay is equal to $\{(1-\alpha)/n\}P_T$.

Now, the instantaneous received power over source to destination link is expressed as

$$P_f = |f|^2 \, \alpha \, P_T \tag{3}$$

Instantaneous received power over source to i^{th} relay link is as:

$$P_{h_i} = |h_i|^2 \, \alpha \, P_T \tag{4}$$

And the instantaneous received power over i^{th} relay to destination link is expressed as

$$P_{g_i} = |g_i|^2 \, \frac{(1-\alpha)}{n} \, P_T \tag{5}$$

By using maximum ratio combining (MRC), different independently faded versions are combined at the destination and total received power at the destination is as

$$P_r = P_f + \sum_{i=1}^{n} \frac{P_{h_i} P_{g_i}}{1 + P_{h_i} + P_{g_i}} \tag{6}$$

The maximum rate of transmission of information per second is the channel capacity per second which analyzes the performance of the system. If the received power is so high due to which noise power is negligible, then channel capacity by Shannon's formula is as:

$$\overline{C}(P_r) = \log_{10}(1 + P_r) \tag{7}$$

After putting the value of P_r from equation (6) in equation (7), we get

$$\overline{C}(P_r) = \log_{10}\left(1 + P_f + \sum_{i=1}^{n} \frac{P_{h_i} P_{g_i}}{1 + P_{h_i} + P_{g_i}}\right) \tag{8}$$

For Nakagami-m distribution, Gamma distributed probability density function (PDF) is expressed as

$$P(P_r) = \frac{1}{\Gamma(m)} \left(\frac{m}{P_{avg}}\right)^m (P_r)^{m-1} e^{\left(\frac{-mP_r}{P_{avg}}\right)} \tag{9}$$

After putting the value of P_r from equation (6) in equation (9), we get PDF as:

$$P(P_r) = \frac{1}{\Gamma(m)} \left(\frac{m}{P_{avg}}\right)^m \left(P_f + \sum_{i=1}^{n} \frac{P_{h_i} P_{g_i}}{1 + P_{h_i} + P_{g_i}}\right)^{m-1} e^{\left(\frac{-mP_r}{P_{avg}}\right)} \tag{10}$$

Here, $\Gamma(.)$ is the gamma function, parameter m is the shape factor and P_{avg} **is the average received power**

over the fading and shadowing effects.
The ergodic capacity analyzes the performance of the system which is the ensamble average of the information rate over the distribution of the elements of the channel which is denoted as

$$\overline{C}_E = \int_0^\infty \overline{C}(P_r) \, P(P_r) \, dP_r \tag{11}$$

or

$$\overline{C_E} = \int_0^\infty \left(\frac{1}{\Gamma(m)} \left(\frac{m}{P_{avg}} \right)^m \left(P_f + \sum_{i=1}^n \frac{P_{h_i} P_{g_i}}{1 + P_{h_i} + P_{g_i}} \right)^{m-1} e^{\left(\frac{-mP_r}{P_{avg}} \right)} \right)$$

$$\log_{10} \left(1 + P_f + \sum_{i=1}^n \frac{P_{h_i} P_{g_i}}{1 + P_{h_i} + P_{g_i}} \right) dP_r$$

(12)

III.   PERFORMANCE ANALYSIS

To analyse the performance of cooperative communication wireless system model, the outage probability of the system also plays an important role in our study.

(a) Outage probability

To analyze the outage behaviour of cooperative communication wireless network, we select a threshold power level (*Pn*) in such a manner that if the total received power is below this level, then the system is said to be in outage for this particular channel and the probability of the system in outage condition is called the outage probability [6],

$$P_{out} = P_{rb} \{ P_r < P_n \} = \int_0^{P_n} P(P_r) dP_r \qquad (13)$$

After putting the value of $P(P_r)$ from eq. (10) in eq.(13), we get

$$P_{out} = \int_0^{P_n} \left(\frac{1}{\Gamma(m)} \left(\frac{m}{P_{avg}} \right)^m \left(P_f + \sum_{i=1}^n \frac{P_{h_i} P_{g_i}}{1 + P_{h_i} + P_{g_i}} \right)^{m-1} e^{\left(\frac{-mP_r}{P_{avg}} \right)} \right) dP_r \qquad (14)$$

Where, P_n is selected threshold power level , m is Nakagami-m parameter, P_{avg} is the average received power

over the fading and shadowing effects and P_r is total power received at the destination as mentioned earlier in

eq.(6).

IV.   SIMULATION RESULTS

In this section, we have shown our results of outage probability, channel capacity and ergodic capacity and total received power versus total average power at the destination. We get the simulated results using MATLAB software.
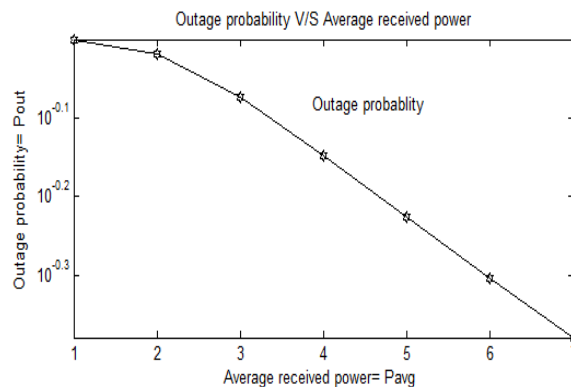
Fig. 4. Outage probability V/S average received power at destination

Fig.4. shows the outage performance at different values of total average power. This curve shows that the outage probability reduces fast as the value of total average power increases causing improvement in the performance of the system. For a particular value of total average power at the destination node, the channel capacity and ergodic capacity increase as shown in fig.5 and fig.6 respectively.
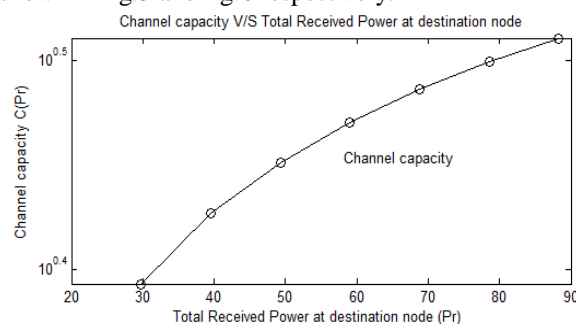
Fig.5. Channel capacity V/S average received power at destination

In this section, we have considered threshold power level P_{th}= 5 dB, Nakagami-m parameter (shape factor) m=2, $P_f = P_{h_i} = P_{g_i} = 20$ dB.
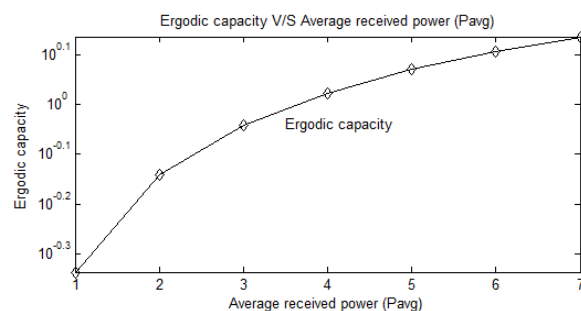
Fig. 6. Ergodic capacity V/S average power received at the destination

As the average value of received power increases, the value of probability density function of Nakagami-m fading distribution decreases resulting in the proper operation of the system due to lower outage probability that is the probability of the system to be in outage state in which state, system does not work properly.

V.   CONCLUSION

From the simulation results, we can observe that system parameters such as channel capacity, ergodic capacity increase and outage probability decreases as the value of average received power increases and it shows that the system becomes more reliable and efficient.

REFERENCES

[1]   A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity-Part I: System description", IEEE Trans. Commun., vol. 51, No. 11, pp. 1927-1938, Nov. 2003.
[2]   A. Sendonaris, E. Erkip, and B. Aazhang, "User Cooperation Diversity-Part II: Implementation aspects and performance analysis", IEEE Trans. Comm., vol. 51, no. 11, pp. 1939-1948, Nov. 2003.
[3]   J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks efficient protocols and outage behaviour", IEEE Trans. Inf. Theory, vol. 50, no. 12, pp. 3062–3080, 2004.
[4]   J. N. Laneman and G. Wornell, "Energy-efficient antenna sharing and relaying for wireless networks", in Proc. Wireless Commun. Networking Conf. 2000, vol. 1, pp. 3062–3080.
[5]   J. N. Laneman and G. Wornell, "Distributed space-time coded protocols for exploiting cooperative diversity in wireless networks", in Proc. Glob. Telecomm. Conf. 2002, vol. 1, pp. 77–81.
[6]   Ankur Bansal and Parul Garg, "Performance Analysis of Coded Cooperation Under Nakagami-m Fading Channels", in Proc. IEEE Int. Conf. on Comm. (ICC), Cape Town, South Africa, May 2010.
[7]   J. N. Laneman, G. W. Wornell, D. N. C. Tse,"An efficient protocol for realizing cooperative diversity in wireless networks", in proc. IEEE Int. Symp. Inform. Theory, Washington, DC, pp. 294, Jun. 2001.
[8]   T. E. Hunter and A. Nostratinia, "Cooperation diversity through coding", in Proc. IEEE Int. Symp. Inform. Theory, Laussane, Switzerland, pp. 220, Jun. 2002

Design and Implementation of a Distributed Database System for the Central Bank of Iraq Using Oracle

Mayson Mohammad Talab[1], Dr. Bassam Ali Mustafa[2]

[1,2][1]Computer Science Dept., Computer Science and Mathematics, University of Mosul, Iraq,

***ABSTRACT**:*

This paper is concerned with design and implementation of a distributed database system for Central Bank of Iraq purposes. Two-tier client/server model has been used to implement the proposed systems architecture which consists of twelve clients spread over office which are connected together through a local area network (LAN) by using three hubs which spread one hub at each floor of office see figure2. Partially replicated databases technique has been used to distribute database. The designed distributed system is homogenous and uses Windows XP operating system and Oracle10g software to implement, connect, and manage the database, whereas user screens and reports have been designed using "Oracle Forms Builder 6i" and "Oracle Reports Builder 6i" respectively. The system has achieved data protection against unauthorized access using combination of passwords and user roles, in addition to providing mechanisms for data loss protection through import and export mechanism. The system has applied in achieving some of the distributed systems goals like transparency, connecting users, and sharing resources.

I. INTRODUCTION

Distributed computing is one of the most recent and important development in the computing era. During recent times we have seen the rapid development of network and data communication technology. With the combination of these two technologies, databases are moving away from a simple centralized model and changing to a decentralized concept to meet the requirements of computing speed, performance, cost effectiveness and increasing the reliability that are not meet by Centralized Computer systems [1][2].Distributed system is a piece of software which ensures that a collection of independent computers that appears to its users as a single coherent system. A distributed database system consists of a collection of sites, each of which maintains a local database system. Each site is able to process local transactions, those transactions that access data only in the single site. In addition, a site may participate in the execution of global transaction, those transactions that access data in several sites. Execution of global transaction requires communication among the sites. Goals of distributed systems**:** Connecting resources and users, distribution transparency, Openness and Scalability [3].Distributed system can be connected physically in a variety of ways such as file server, peer-to-peer model and client-server architecture. In client server model, server does most of the data management work. This means that all query processing, transaction management and storage management is done at the server. Client server has many model such as two-tier client/server model which is simplest client/server architecture, also called multiple-client/single-server approach where database is stored on only one server. Processing is split between the client and the server, and there is much less data traffic on the network. In three-tier client/server some databases can be stored on a client PC's own hard drive while other databases that the client might access are stored on the LAN's server. In multi-tier approach, the client PCs and servers known as application servers, and other servers known as database servers. It is more sophisticated client/server architecture where there are multiple serves in the system (the so-called multiple-client/multiple-server approach). There are three types To communicate between server and its clients which are Remote Procedure Calls (RPC), Remote Data Access (RDA) and Queued Message Processing (QMP) [3][4][5] .

II. ANALYSIS AND DESIGN OF DISTRIBUTED SYSTEM

2.1.Requirements Analysis: Before designing any system requirements of the work should be analyzed .Therefore, requirements analysis are a very important step in order to form a complete idea of the system to be designed. Methods used in collecting data are Personal interview, questionnaire or referendum, observation, record Searching and appreciation Estimation. During the process of

analyzing the system and knowing reality of work in the Iraqi Central Bank ,it was concluded the following matters: First: -CBI includes the following departments: (Department of Human Resource Management, Department of Public Accounts, Department of Current Accounts, Department of Issuance ,Department of Computer, Department of Internal Supervision, foreign exchange Department, Department of Banking, Department of Legal).

Second: - systems that are currently working in the bank is a group of independent systems in different languages are as follows:

1. FoxPro language system manages work of departments of current accounts, public accounts and issuance via LAN network.
2. Integrated Central Bank System(ICBS) is purchased from Lebanese company and it is under test yet .It does not meet requirements of the Bank's work exactly because it has been designed according to the general world standards system . also it is very expensive because it is just executive system does not contain programs and any update requires huge amounts of money and travel to Lebanon. Finally it adds complexity and delay of the daily work
3. Microsoft Access system to regulate reports to clearing house system.

   The multiplicity of systems and independence has led to confusion and duplication of work daily as following:
1. after end of the work of clearing house, credit and debit documents are configured for each participant bank and then handed to the computer employee to be entered to the FoxPro system and also to the ICBS
2. forms of central deposit check is entered to the FoxPro system in two stages :the first one is entered account number and form number and the amount of the form and the second stage is entered its checks at different screen.
3. forms of bank deposit check is entered in stages as: Group of checks are sorted manually according to banks and are entered to Microsoft Access system and print a report for each bank , then printer employee prints debit and credit forms for each bank based on reports of Microsoft Access System , after that handing these forms to computer employee to be entered to the FoxPro system to update balances of banks, than the completion of all forms are delivered with the credit or debit form in addition to the Access report to a representative of the bank
4. the same previous steps are entered to ICBS .
5. final total of form of the deposit check does not appear at monthly reports which are printed by ICBS .
6. At FoxPro system, dual forms between current accounts and public accounts is entered in two stages: First:- employee of current accounts enters its own entity. Second:- employee of public accounts enters its own entity.
7. At FoxPro system ,current accounts system has two copies of system in which one for Iraqi dinar currency and other copy for U.S. dollar currency
8. Complexity of the input of ICBS compared with FoxPro system, for example, The ICBS account number consists of 16 digits either FoxPro consists of 6 digits only.
9. At ICBS, data entry process is slow in spite of input speed feature of the Oracle.
10. At FoxPro system, emergence of either increase or decrease fils of the final balance in monthly reports of current accounts .
11. At ICBS, there is no final summary report for the CBI work which final totals include the work of current accounts and public accounts and the Fund in order to match the movements and ensure the authenticity of the daily work.

The goal of this paper eliminates disadvantage of these systems and groups their advantages in on complete distributed system for Iraqi Central Bank.

2.2.Database Design:we use Top-down design that is started by identifying the data sets, and then defining the data elements for each of those sets. This process involves the identification of different entity types and the definition of each entity's attributes. Database design usually splits the design phase into two parts logical and physical design.

2.2.1.Logical Design: It shows a general view of the structure of the system and its basic components. Results of logical design include series of charts and data.

2.2.1.1.Specifying Entities and Their Attributes: After analyzing requirements of the work**,** we identify the following: (17)entities for current accounts system ,(4) entities for public accounts system, (5) entities for fund system, (5) entities for treasury system, (3) entities for vacations system, (5) entities for individual issues system, (4) entities for Security system, (2) entities for clearing house system and (7) entities for salary system. Here we show entities of current accounts system as:

1- Iraqi balance entity (CACMAS):It contains all information of accounts and the balances with Iraqi dinar currency. It is essential entity and contains account number as primary key(PK).It includes following attributes:(account number, account name, global account number, account state, account nature, account open date, open balance, current balance, account type, branch number, currency type).

2- Dollar balance entity (DCACMAS):It is similar to attributes of Iraqi balance entity (CACMAS) ,but it is for U.S. Dollar currency

3- Iraqi daily movements entity (CACTRAN):It Includes all types of movements which are entered during daily working for the Iraqi dinar currency. It contains account number as foreign key(FK). It includes following attributes:(account no., check no., movement date, movement nature, office no., movement amount, check document no., currency type, document no., branch no., movement type).

4- Dollar daily movements entity (DCACTRAN): It is similar to attributes of Iraqi daily movements entity (CACTRAN),but it is for U.S. Dollar currency

5- Iraqi historical movements entity(CACHIST): It Includes all types of movements which are entered during current year for Iraqi dinar currency. It contains account number as foreign key(FK).It's attributes are similar to Iraqi daily movements entity (CACTRAN).

6- Dollar historical movements entity(DCACHIST): It is similar to attributes of Iraqi historical movements entity(CACHIST),but it is for U.S. Dollar currency

7- Iraqi old historical movements entity(OLD_CACHIST): It Includes all types of movements which are entered during past years for Iraqi dinar currency. It contains account number as foreign key(FK It's attributes are similar to Iraqi daily movements entity (CACTRAN).

8- Dollar old historical movements entity(OLD_DCACHIST): It is similar to attributes of Dollar old historical movements entity (OLD_DCACHIST), but it is for U.S. Dollar currency

9- Iraqi total accounts entity (CACGLOBAL):It includes total accounts numbers for Iraqi dinar currency according to type of accounts. It contains global account number as primary key(PK). It includes following attributes:(global account no., global account name).

10- Dollar total accounts entity (DCACGLOBAL):It includes total accounts numbers for U.S. dollar currency according to type of accounts. It contains global account number as primary key(PK). It includes following attributes:(global account no., global name).

11- Check books entity(CHEKBOOK): It includes information of check books which are given to banks. It contains account number as foreign key(FK). It includes following attributes:(account no., book no., first check no., last check no.).

12- Suspended checks entity(STOPCHEK): It includes information of checks which are prevented from drawing for different reasons. It contains account number as foreign key (FK). It includes following attributes:(account no., check no., suspended date).

13- Exchange price entity (PAY_RATE): It includes daily price of exchange U.S. dollar currency according to Iraqi dinar currency. It includes following attributes:(pay price, pay date).

14- Movement type entity(TRAN_NAM) : It includes code and name of each type of movement. It contains code as primary key(PK). It includes following attributes:(movement code, movement name).

15- Branch type entity(TYP_NAM): It includes types of all current accounts. It contains code as primary key(PK). It includes following attributes:(account code, account name).

16- Debit form entity (DOC_NUM): It includes information of debit forms which are daily given to each bank after end of work of clearing house. It contains current account as foreign key(FK). It includes following attributes:(account no., document no., document date, numeric number, written number).

17- Branches names(MSRF_FR):It includes odes and names of all banks branches. It contains primary key(PK). It includes following attributes:(branch code, branch name).

2.2.1.2. Structures of Entities: We show structure of Iraqi balance entity (CACMAS) and Iraqi daily movements entity (CACTRAN) as example.

Table 1:Table of Iraqi Balance Entity (CACMAS).

Name of attribute	Programmed name	Type	Length	Key
Account number	ACCNO	NUMBER	7	PK
Account name	NAME	CHAR	75	
Global account number	GLOBACC	NUMBER	7	
Account state	ACCSTAT	NUMBER	1	
Account nature	ACCNAT	NUMBER	1	
Account open date	STARTDATE	DATE	8	
Open balance	OPENBAL	NUMBER	(30,3)	
Current balance	CURNTBAL	NUMBER	(30,3)	
Account type	ACCTYPE	NUMBER	3	
Branch number	BRNO	NUMBER	5	
Current type	MONYTYPE	NUMBER	3	

Table 2: Table of Iraqi Daily Movements Entity (CACTRAN)

Name of attribute	Programmed name	Type	Length	Key
Account name	ACCNO	NUMBER	7	FK
Movement number	TRAN	NUMBER	10	
Movement date	DOCDATE	DATE	8	
Movement type	TRANCODE	NUMBER	1	
Office account no.	ACCNO2	NUMBER	7	
Movement amount	AMOUNT	NUMBER	(30,3)	
Check document no.	DOCNO2	NUMBER	10	
Movement type	TYP	NUMBER	3	
Currency type	MONYTYPE	NUMBER	3	
Movement no.	DOCNO	NUMBER	10	
Branch no.	FR	NUMBER	3	

2.2.1.3 **Relations Between Tables:** Type of used relation is one to many(1:M). There are relations between current account entities as: Primary key (account no.) of Iraqi balance entity (CACMAS) is connected to foreign keys (account numbers) of following entities:(Iraqi daily movements entity (CACTRAN), Iraqi historical movements entity(CACHIST), Iraqi old historical movements entity(OLD_CACHIST), Check books entity(CHEKBOOK), Suspended checks entity(STOPCHEK), Debit form entity (DOC_NUM)) . Primary key(global account code) of Iraqi total accounts entity (CACGLOBAL) is connected to foreign key (global account code) of Iraqi balance entity (CACMAS). Primary key(account code) of Branch type entity(TYP_NAM) is connected to foreign key(account type) of Iraqi balance entity (CACMAS). Primary key(movement code) is connected to foreign keys(movement type) of following entities:(Iraqi daily movements entity (CACTRAN), Iraqi historical movements entity(CACHIST). Primary key(account cod) of Branches names(MSRF_FR) is connected to foreign key(branch no.) of Iraqi daily movements entity (CACTRAN). Primary key (account no.) of Dollar balance entity (DCACMAS) is connected to

foreign keys (account numbers) of following entities:(Dollar daily movements entity (DCACTRAN), Dollar historical movements entity(DCACHIST), Dollar old historical movements entity(OLD_DCACHIST), Check books entity(CHEKBOOK), Suspended checks entity(STOPCHEK)) . Primary key(global account code) of Dollar total accounts entity (DCACGLOBAL) is connected to foreign key (global account code) of Dollar balance entity (DCACMAS). Primary key(account code) of Branch type entity(TYP_NAM) is connected to foreign key(account type) of Dollar balance entity (DCACMAS).

2.2.1.4.Relationship Modeling: We model entities and their relationships as show in bellow Fig(1).
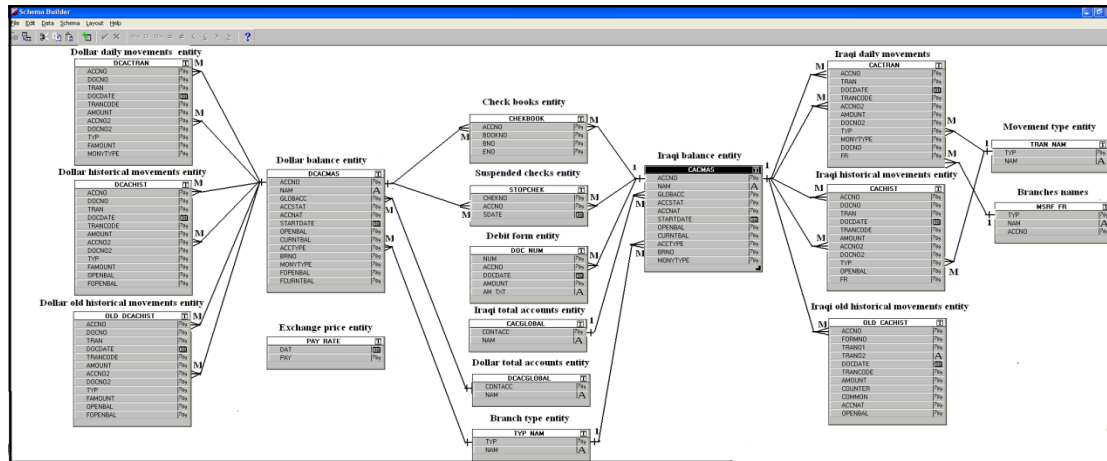
Figure1: Modeling of Entities and Relationships of Current Account System

*2.2.1.5.***Method of Distribution**: Finally decide on the method of distribution where we use replication mechanism for partially replicate tables of vacations, individual issues, clearing house, treasury ,fund and salary systems in which we have copy of tables on server and local clients.

*2.2.2.***Physical Design**: After logical design we determine the physical structure of the database; in order to enable computer programmers to create programs to design a working distributed system.
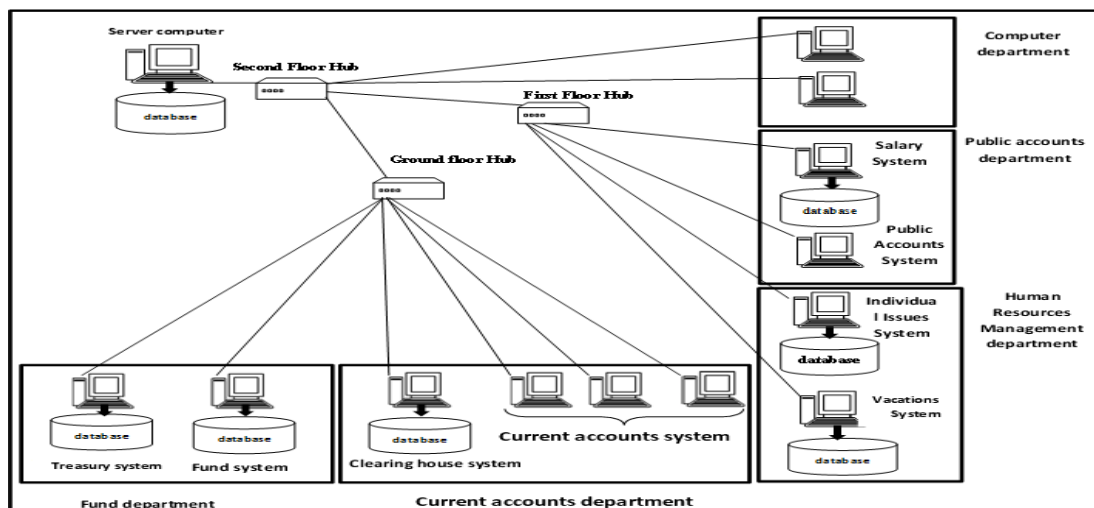
Figure2:Architecture of proposed distributed system of central bank

Figure(4) show the structure of proposed distributed system for Iraqi Central Bank. We need to three hubs, the first one is considered as a central hub because we use star topology to connect Iraqi Central Bank's LAN network. Central hub is located at second floor which serves second hub at first

floor and third hub at ground floor. Also we need server computer to serve distributed system which contain global database called "BANK_DB" that contains (55) tables of all Central Bank systems. Finally we need twelve computer devices which six computer of them considered as clients without local databases for current account department, public account department, and computer department. Where other six clients computer considered as clients with local data base for human resource management department, clearing house, treasury system, fund system, salary system. After preparing all devices of distributed systems we must configure these devices as: We install server computer by Windows Server 2003,Oracle database 10g and Oracle Developer 6i then create administrator user to build all tables of systems. Also we identify IP , host name, workgroup for server. Whereas clients computer is installed by Window XP , Oracle Developer 6i,and Oracle database 10g only for clients should hold local database. Then we configure listener at server to listen for all clients order, after we sure that connection inside network is proper work we create databases links which use to connect between distributed databases on clients and global database "BANK_DN" on server .Then we build applications of all systems depending on Oracle Forms Builder which contains Triggers and Program Units that help programmers to build all necessary codes and depending on Oracle Reports Builder to build all required repots.

III.   IMPLEMENTATION OF THE SYSTEM:

At this step, we execute distributed system of Iraqi Central Bank from any client connect with proposed distributed system network by open icon called "Central Bank of Iraq", then we enter name and code privileged user which is identified by security system to give high security for all systems.
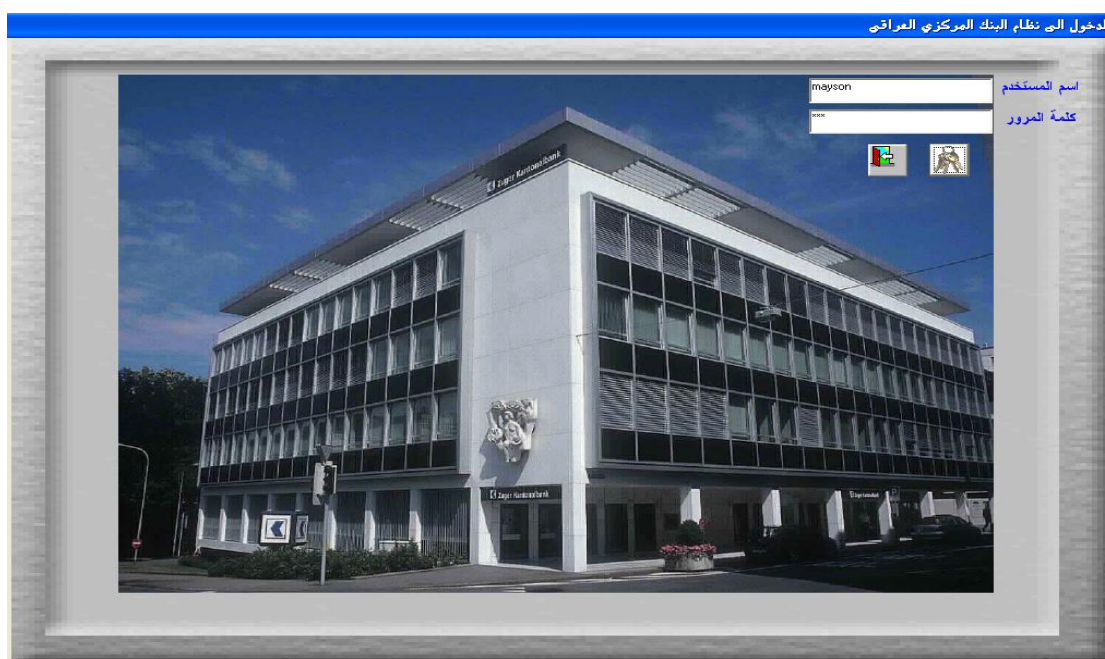
Figure3:Entering form of Iraqi central bank

If name and code of user is true then forms in Fig(4) appears to enable user entering to required system as his privilege.

*Figure4:*Main form of Iraqi central bank

IV.   Maintains of the System:

This process continues throughout the life of the system to make sure of proper working as well as modifying the system to align it with the new requirements of the work environment, such as changing laws, regulations and business rules.

V.   Export/Import:

The Oracle export and import utilities are the most primitive method of sharing data among databases and are also used as part of a backup and recovery strategy. Import is the utility that reads this file and executes the SQL statements to re-create the objects and populate tables. A full database export creates a file that you can use to re-create the entire database.Figure() show how we are programed export and import technique
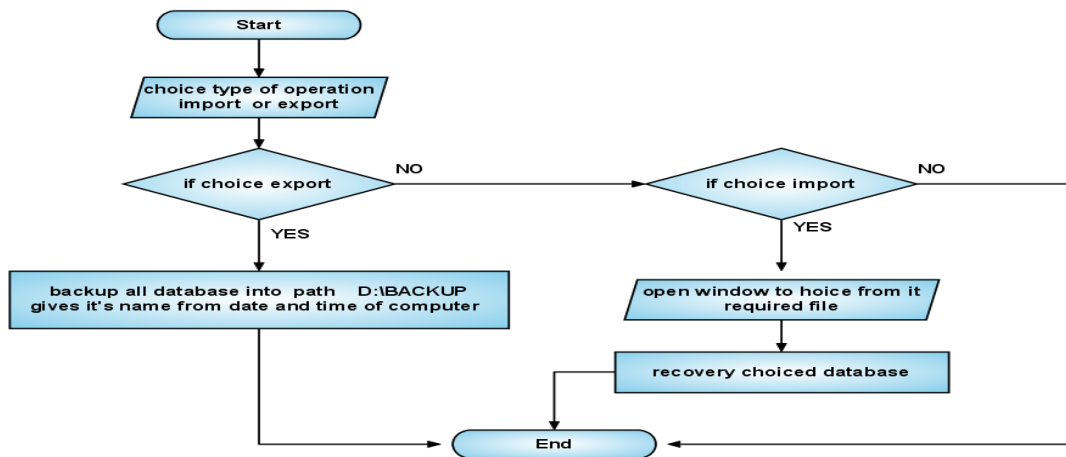
figure(5) import and export of Iraqi central bank

VI.   CONCLUSIONS

After designing, implementing and applicating a distributed database system in the Central Bank of Iraq, we conclude that the analysis of the existing system in the organization before you begin the process of system design and move to right step from logical to physical design within the structure of the system has had a perfect impact on the proper and effective representation of databases. The proposed system has succeeded to achieve its own objectives of design and representation of distributed database for Central Bank and distribution of tasks on more than one site which led to the speed of work and reduce the burden of processing, using Oracle and its software accessories because its possession of specialized tools for this purpose. Also it has high security through a privilege system that can be grouped in the form of roles granted to users smoothly as well as the application distribution technologies such as replication technology. System has succeeded in achieving the basic concepts of distributed systems such as transparency, connecting users and resources with each other, as

well asits ability for daily backup of the database. Using a trial period for the system helped to limit errors and added new requirements of users that occurred during the work and correcting path of the system to accept different views of the beneficiaries to become a computer system acceptable to work where it has achieved data sharing among the different units, and facilitated follow-up daily business and access to information , reports, and periodic extraction of statistics easily and quickly, which helped to save time and effort

VII.   RECOMMENDATIONS

The search resulted in a set of recommendations for the work which can be implemented in the future as follows:

1 - Extend the distributed system to connect branches of the Central Bank of Iraq, which are branch of Mosul, Basra, Erbil, Sulaymaniyah and Bagdad via WAN network.
2 – Extend distributed system to connect service of clearing house system of the Central Bank of Iraq with all existing banks via WAN network.
3 - Design and programming distributed system to control on banking via Internet, which allows Central Bank of Iraq to follow the banks works through a central control system where database is stored in the CBI.

REFERENCES

[1]   Alyaseri ,S., "Distributed University Registration Database System Using
[2]   Oracle9i",College of engineering, University of Basrah, 2010.
[3]   Newton ,M., "Aspects of Distributed Databases" , Computer Science,2012.
[4]   Anderw, S., Steen ,M. V.,"DISTRIBUTED SYSTEMS PRINCIPLES AND
[5]   PARADIGAMS", VRIJE UNIVERSITEITAMSTERDAM, The Netherlands, RENTICE
[6]   HALL, 2002.
[7]   Coulouris, G., Kindberg,T.,Dollimore,J.,Blair,G. ,"Distributed Systems: Concepts and
[8]   Design". Fifth Edition, Addison Wesely.",Oracle,2012.
[9]   Özsu, M. T., Valduriez, P., " Priciples of Distributed Database Systems". Third   Edition,
[10]   Business Media LLC, 2011.

Quantitative Treatment of Hiv/Aids Inthe Human Micro-Vascular Circulating Blood System

[1]Enaibe A. Edison, [2]Osafile E. Omosede , [3]John O. A. Idiodi

[1,]*Department of Physics, Federal University of Petroleum Resources*
P.M.B. 1221 Effurun, Nigeria.
[2] *Department of Physics, University of Benin, Benin City, Edo State, Nigeria.*

ABSTRACT

Vibration is the cause ofall that exists.Generally, any form of matterincluding biological systemscould be described by vibration and hence represented by a purely defined wave form. This study presents a model for determining the dynamic mechanical characteristics of HIV/AIDS in the human blood circulating system. Our work assumes that the physical dynamic components of the HIV responsible for their destructive tendency are $b\,\lambda$, $n'\lambda$, $\varepsilon'\,\lambda$ and $k'\,\lambda$ been influenced by the multiplicative factor λ whose physical range of interest is $0 \le \lambda \le 13070$. We constructed the constitutive carrier wave equation on the basis of the vibratory dynamic components of the human (host) parameters and those of the HIV (parasite). It is established in this study that when the HIV enters the human blood circulating system, it takes between 60 to 240 days before its absolute effects would begin to manifest. This in the literature of clinical disease is referred to as the window period. The negative influence of the HIV in the human system becomes intense and more pronounced when the HIV is about 5 years (60 months) counting from the day it is contacted. This study revealed that AIDS actually results when these destructive dynamic components of the HIV gradually become equal to their corresponding active dynamic components in the human blood circulating system. This is when the range of the multiplicative factor $12803 \le \lambda \le 13070$. The time it takes the HIV infection to degenerate into AIDS if uncontrolled is about 8 years (96months). The displacement of the carrier wave that describes the biological system of Man finally goes to zero - a phenomenon called death, when the multiplicative factor approaches the critical value of $13070$and the time it takes to attain this value is about 11 years (132 months).

KEYWORDS: *latent vibration, 'host wave', 'parasitic wave', carrier wave, HIV/AIDS, amplitude, oscillating phase.*

I. INTRODUCTION

Some waves in nature behave parasitically when they interfere with another one. Such waves as the name implies has the ability of transforming the initial characteristics and behaviour of the interfered wave to its own form and quality after a given period of time. Under this circumstance, all the active constituents of the interfered wave would have been completely eroded and the resulting wave which is now parasitically monochromatic, will eventually attenuateto zero, since the 'parasitic wave' does not have its own independent parameters for sustaining a continuous existence.

The role of Human-Immunodeficiency Virus (HIV) in the normal blood circulating system of man (host) has in general been poorly understood. However, its role in clinical disease has attracted increasing interest. Human immunodeficiency virus (HIV) infection / acquired immunodeficiency syndrome (AIDS) is a disease of the human immune system caused by HIV [1].During the initial infection a person may experience a brief period of influenza-like illness. This is typically followed by a prolonged period without symptoms. As the illness progresses it interferes more and more with the immune system, making people much more likely to get infections, including opportunistic infections, which do not usually affect people with immune systems. In the absence of specific treatment, around half of the people infected with HIV develop AIDS within ten years [2] and average survival time after infection with HIV is estimated to be 9 to 11 years [3].

According to the literature of clinical diseases, the HIV feeds on and in the process kills the active cells that make up the immune system. This is a very correct statement but not a unique understanding. There is also a cause (vibration) that gives the HIV its own intrinsic characteristics, activity and existence. It is not the Human system that gives the HIV its life and existence, since the HIV itself is a living organism and with its own peculiar characteristics even before it entered the system of Man.

It is the vibration of the unknown force that causes life and existence. Therefore, for any active matter to exist it must possess vibration. The human heart stands as a transducer of this vibration. Fortunately the blood stands as a means of conveying this vibration to all units of the human system.The cyclic heart contraction generates pulsatile blood flow and latent vibration. The latent vibration is sinusoidal and central in character, that is, it flows along the middle of the vascular blood vessels. It orients the active particles of the blood and sets them into oscillating motion with a unified frequency as it passes. Elasticity of the vascular blood vessels supports pulsatile blood flow, connectivity network of the blood circulating system, and not the latent vibration. Man and the Human-Immunodeficiency Virus (HIV) are both active matter, as a result, they must have independent peculiar vibrations in order to exist. It is the vibration of the HIV that interferes with the vibration of Man (host) in the blood circulating system after infection. The resultant interference of the vibration is parasitically destructive and it slows down or makes the biological system of Man to malfunction since the basic intrinsic parameters of the resident wave function have been altered.

The activity of the HIV is everywhere the same within the human blood circulating system, mutation if at all does not affect its activity. That the HIV kills slowly with time shows that the wave functions of the HIV and that of the host were initially incoherent. As a result, the amplitude, angular frequency, wave number and phase angle of the host which are the basic parameters of vibration were initially greater than those of the HIV. The human aorta is the main truck of a series of vessels which convey the oxygenated blood from the heart to the tissues [4]. It is described in several portions, viz, the ascending aorta the arch of the aorta, and the descending aorta. The ascending aorta is about 7cm (0.07m) in length and it has a radius of 1.5cm (0.015m). Arch of the Aorta is about 1.8cm (0.018m) in length and its radius is 1cm (0.01m) while the descending aorta has a length of 1.14cm (0.0114m) and a radius of 1.11cm (0.0111m). The human artery is an extension of the aorta and there are various forms with approximate radius of about 0.4cm (0.004m) [5]. The smallest vessels, the capillaries, have a diameter of about 5×10^{-6} m and 10×10^{-6} m, so that the red blood cells whose diameter is about 8×10^{-6}mcan pass through it [6]. There are about 250 capillaries/mm^{3}of body tissues and average length of a capillary is about 600 microns (600×10^{-6}m). However, we are going to utilize only the ascending parameters of the aorta in our calculation and assume a uniform geometry and structure for all the vascular blood vessels. This assumption is reasonable since the latent

vibration takes its first unique course through the ascending aorta. The human veins are not taken into consideration in our work, because it only conveys denatured blood (deficient in oxygen and food nutrients) to the human heart for reactivation. Also because of the limited length of the human capillaries the exchange process of active blood in this region of space does not take time as a result our computation will not include the capillaries as well. Human blood is a liquid tissue composed of roughly 55% fluid plasma and 45% cells. The three main types of cells in blood are red blood cells, white blood cells and platelets. 92% of blood plasma is composed of water and the other 8% is composed of proteins, metabolites and ions [7]. The density of blood plasma is approximately 1025 kg/m^3 and the density of blood cells circulating in the blood is approximately 1125 kg/m^3. Blood plasma and its contents are known as whole blood [8]. The average density of whole blood for a human is about 1050 kg/m^3. Blood viscosity is a measure of the resistance of blood to flow, which is being deformed by either shear or extensional strain [9]. The dynamic viscosity (μ) of the human blood at 37^{0}C is

usually between 0.003kgm^{-1}s^{-1}and0.004kgm^{-1}s^{-1}, the arterial blood perfusion rate (w_b) is 0.5kgm^{-3}s^{-1}[10]. The viscosity of blood thus depends on the viscosity of the plasma, in combination with the particles. However, plasma can be considered as a Newtonian fluid, but blood cannot due to the particles which add non-idealities to the fluid.If a wave is to travel through a medium such as water, air, steel, or a stretched string, it must cause the particles of that medium to oscillate as it passes [11]. For that to happen, the medium must possess both mass (so that there can be kinetic energy) and elasticity (so that there can be potential energy). Thus, the medium's mass and elasticity property determines how fast the wave can travel in the medium.The initial characteristics of a given wave with a definite origin or source can best be determined by the use of a sine wave function. However, for the deductive determination of the initial behaviour of a wave whose origin is not certain, the cosine wave function can best be effectively utilized.

The principle of superposition of wave states that if any medium is disturbed simultaneously by a number of disturbances, then the instantaneous displacement will be given by the vector sum of the disturbance which would have been produced by the individual waves separately. Superposition helps in the handling of complicated wave motions. It is applicable to electromagnetic waves and elastic waves in a deformed medium provided Hooke's law is obeyed Generally, it is the human blood that responds to the latent vibration from the heart with a specified wave form. The blood then propagates away from the region of the disturbance and in the process circulates oxygen and food nutrients to nourish the biological cells of the human system. Any alteration to this process results to starvation, a gradual weakening of the fundamental cells and a subsequent breakdown of the entire human system if uncontrolled.

II.   RESEARCH METHODOLOGY

[1]   The wave characteristics of blood in the circulating system of a normal person free from HIV/AIDS is assumed to be measured and the following observations about the wave function were recorded: (i) the amplitude, a (ii) the phase angle, ε (iii) the angular frequency, n and(iv) the wave number, k. Note that a, ε and n are assumed to be constant with time in the human system except for some fluctuating factors, e.g. illness, which of course can only alter them slightly and temporarily.

[2]   The wave characteristics of blood in the circulating system of a HIV/AIDS infected candidate, whose immune count rate is already zero is assumed to be measured. The following observations about the wave properties were recorded: (i) the amplitude, b (ii) the phase angle, ε' (iii) the angular frequency, n' (iv) the wave number, k'.

[3]   Since the immune system of the HIV/AIDS individual is exactly zero, the measured wave function shall depend entirely on the vibrating property of the HIV only as every other active constituents of the blood system have been completely eroded. That is, the measured wave characteristic of the two candidates cannot be the same and whatever that makes the difference are the attributes of the HIV.

[4]   The measured wave properties of the HIV infected candidate is independent of intrinsic variables such as the number, size, mass and of course the mutation property (if at all) of the invading HIV.

[5]   The measured wave characteristic of the HIV infected candidate is the same everywhere within the host (Man). That is, irrespective of the occupation of the HIV in the host system, be it in the capillary, vein, liver, bone marrow, or in the brain, the activity of the HIV is the same everywhere and hence the wave function must be the same. Of course the veins only carry denatured blood back to the heart for activation and the nutrient blood takes negligible time in the capillary during the energy exchange process.

[6]   The wave properties of the HIV cannot be directly measured since it does not have independent existence outside the host system. As a result, the wave function of the HIV can only be deductively measured.

[7]   If the HIV exists it must have a peculiar vibration of its own which must be independent of the vibration of the host. If the attributes of this vibration are known, then, it can be selectively destroyed from the body of the host.

[8]   That the HIV kills slowly with time shows that the wave function of the HIV and that of the human (host) system are incoherent. As a result, the amplitude, frequency and the phase angle of vibration of the host were initially greater than those of the vibration of the HIV (parasite).

The aim of this work is to describe the biomechanics of HIV/AIDS and to report the methodology developed in our laboratory to characterize the dynamics of the 'host wave' and those of 'HIV wave' in the constitutive carrier wave equation propagating in the human blood circulating system. Understanding wave propagation in arterial walls, local hemodynamics, and wall shear stress gradient is important in understanding the mechanisms of cardiovascular function. Arterial walls are anisotropic and heterogeneous, composed of layers with different bio-mechanical characteristics which make the understanding of the mechanical influences that arteries contribute to blood flow very difficult [12].

This paper is outlined as follows. Section 1, illustrates the basic concept of the work under study. The mathematical theory is presented in section 2. We present the results obtained in section 3. While in section 4, we present the analytical discussion of the results obtained. The conclusion of this work is shown in section 5, and this is immediately followed by an appendix and a list of references.

$$y = \sqrt{\left(a^2 - b^2\lambda^2\right) - 2\left(a - b\lambda\right)^2 \cos\left((n - n'\lambda)t - (\varepsilon - \varepsilon'\lambda)\right)} \times \cos\left((\vec{k} - \vec{k}'\lambda).\vec{r} - (n - n'\lambda)t - \left|\partial E\right|\right) \quad (2.29)$$

where we have redefined the resultant amplitude A as

$$A = \sqrt{\left(a^2 - b^2\lambda^2\right) - 2\left(a - b\lambda\right)^2 \cos\left((n - n'\lambda)t - (\varepsilon - \varepsilon'\lambda)\right)} \quad (2.30)$$

The spatial oscillating phase and the total phase angle E are respectively given by

$$\phi = \cos\left((\vec{k} - \vec{k}'\lambda).\vec{r} - (n - n'\lambda)t - \left|\partial E\right|\right) \quad (2.31)$$

$$E = \tan^{-1}\left(\frac{a\sin\varepsilon - b\lambda\sin((n - n'\lambda)t - \varepsilon'\lambda)}{a\cos\varepsilon - b\lambda\cos((n - n'\lambda)t - \varepsilon'\lambda)}\right) \quad (2.32)$$

By definition: the modulation angular frequency is given by $(n - n'\lambda)$, the modulation propagation constant is $(\vec{k} - \vec{k}'\lambda)$, the phase difference δ between the two interfering waves is $(\varepsilon - \varepsilon'\lambda)$, the interference term is given by ; $2\left(a - b\lambda\right)^2 \cos\left((n - n'\lambda)t - (\varepsilon - \varepsilon'\lambda)\right)$, while waves out of phase interfere destructively according to $\left(a - b\lambda\right)^2$ and waves in-phase interfere constructively according to $\left(a + b\lambda\right)^2$. In the regions where the amplitude of the carrier wave is greater than either of the amplitude of the individual wave, we have constructive interference that means the path difference is $(\varepsilon + \varepsilon'\lambda)$, otherwise, it is destructive in which case the path difference is $(\varepsilon - \varepsilon'\lambda)$. If $n \approx n'$, then the average angular frequency say $(n + n')/2$ will be much more greater than the modulation angular frequency say $(n - n')/2$ and once this is achieved, then we will have a slowly varying carrier wave with a rapidly oscillating phase.

The total differentiation of the total phase angle E gives the exact total phase angle. That is

$$\partial E = \frac{\partial E}{\partial \varepsilon}d\varepsilon + \frac{\partial E}{\partial \varepsilon'}d\varepsilon' \quad (2.33)$$

$$\left|\partial E\right| = \sqrt{\left(\frac{\partial E}{\partial \varepsilon}\right)^2 + \left(\frac{\partial E}{\partial \varepsilon'}\right)^2} \quad (2.34)$$

There is need for us to make the total phase angle of the oscillating phaseof the carrier wave not to depend explicitly on time, since this would enhance quality results so that the subsequent wave form does not produce irregular complex behaviour.

$$\frac{\partial E}{\partial \varepsilon} = \frac{a^2 - ab\lambda\cos\left((\varepsilon + \varepsilon'\lambda) - (n - n'\lambda)t\right)}{a^2 + b^2\lambda^2 - 2ab\lambda\cos\left((\varepsilon + \varepsilon'\lambda) - (n - n'\lambda)t\right)} \quad (2.35)$$

$$\frac{\partial E}{\partial \varepsilon'} = \frac{-b^2\lambda^3 + ab\lambda^2\cos\left((\varepsilon + \varepsilon'\lambda) - (n - n'\lambda)t\right)}{a^2 + b^2\lambda^2 - 2ab\lambda\cos\left((\varepsilon + \varepsilon'\lambda) - (n - n'\lambda)t\right)} \quad (2.36)$$

Equation (2.29) is now the constitutive carrier wave equation necessary for our study. It describes the activity and performance of most physically active systems. As the equation stands, it is a carrier wave, in which it is only the variation in the intrinsic parameters of the 'parasitic wave' that determines the life span of the active biological system which it describes.

2.1 Equation of motion of the carrier wave in the blood vessels of the host

The carrier wave given by (2.29) can only have a maximum value if the spatial oscillating phase is equal to 1. Hence

$$y_m = \sqrt{\left(a^2 - b^2\lambda^2\right) - 2\left(a - b\lambda\right)^2 \cos\left((n - n'\lambda)t - (\varepsilon - \varepsilon'\lambda)\right)} \quad (2.37)$$

$$\frac{d\,y_m}{dt} = (n - n'\lambda)(a - b\lambda)^2 \sin\left((n - n'\lambda)t - (\varepsilon - \varepsilon'\lambda)\right) \times \left((a^2 - b^2\lambda^2) - 2(a - b\lambda)^2 \cos\left((n - n'\lambda)t - (\varepsilon - \varepsilon'\lambda)\right)\right)^{-\frac{1}{2}}$$

(2.38)

$$\frac{d^2 y_m}{dt^2} = (n - n'\lambda)^2 (a - b\lambda)^2 \cos\left((n - n'\lambda)t - (\varepsilon - \varepsilon'\lambda)\right) \times \left((a^2 - b^2\lambda^2) - 2(a - b\lambda)^2 \cos\left((n - n'\lambda)t - (\varepsilon - \varepsilon'\lambda)\right)\right)^{-\frac{1}{2}}$$

$$-\,(n - n'\lambda)^2 (a - b\lambda)^4 \sin^2\left((n - n'\lambda)t - (\varepsilon - \varepsilon'\lambda)\right) \times \left((a^2 - b^2\lambda^2) - 2(a - b\lambda)^2 \cos\left((n - n'\lambda)t - (\varepsilon - \varepsilon'\lambda)\right)\right)^{-\frac{3}{2}}$$

(2.39)

The equation of motion obeyed by the constitutive carrier wave as it propagates along the human blood vessels experiences two major resistive factors. Firstly, the resistance posed by the elasticity of the walls of the blood vessels and secondly, the elastic resistance of the blood medium. The medium's mass and elasticity property determines how fast the carrier wave can travel in the medium. Consequently, the equation of motion would be partly Newtonian due to the fluidize nature of the blood medium and non-Newtonian due to the particle constituent of the blood which creates nonidealities. We can therefore write the equation of motion as

$$F = -\mu\left(\frac{dy^2}{dt}\right) - \sigma\,y^2$$

(2.40)

$$\rho V \frac{d^2 y}{dt^2} + 2\mu\,y\left(\frac{dy}{dt}\right) + \sigma\,y^2 = 0$$

(2.41)

Where ρ is the density of the human blood (kgm-3), V is the volume of the blood vessel which is considered to be cylindrical vascular geometry ($\pi r^2 l$) and the unit is (m3), μ is the dynamic viscosity of blood (kgm-1s-1) and σ is the elasticity of the blood medium. The influence of gravity on the flow of blood is assumed to be negligible. Hence for maximum value of the carrier wave we then rewrite (2.41) as

$$\rho V \frac{d^2 y_m}{dt^2} + 2\mu\,y_m\left(\frac{dy_m}{dt}\right) + \sigma\,y_m^2 = 0$$

(2.42)

Now with the following boundary conditions that at time $t = 0$, $\lambda = 0$ in (2.37) – (2.39) we obtain

$$y_m = \left(a^2 - 2a^2 \cos(-\varepsilon)\right)^{\frac{1}{2}} \quad (2.43)$$

$$y_m^2 = \left(a^2 - 2a^2 \cos(-\varepsilon)\right) \quad (2.44)$$

$$\frac{dy_m}{dt} = n\,a^2 \sin(-\varepsilon)\left(a^2 - 2a^2 \cos(-\varepsilon)\right)^{-\frac{1}{2}} \quad (2.45)$$

$$\frac{d^2 y_m}{dt^2} = n^2\,a^2 \cos(-\varepsilon)\left(a^2 - 2a^2 \cos(-\varepsilon)\right)^{-\frac{1}{2}} - n^2\,a^4 \sin^2(-\varepsilon)\left(a^2 - 2a^2 \cos(-\varepsilon)\right)^{-\frac{3}{2}} \quad (2.46)$$

When we substitute (2.43) - (2.46) into (2.42) we get

$$\rho V \left(\frac{n^2 a^2 \cos(-\varepsilon)}{(a^2 - 2a^2 \cos(-\varepsilon))^{1/2}} - \frac{n^2 a^4 \sin^2(-\varepsilon)}{(a^2 - 2a^2 \cos(-\varepsilon))^{3/2}} \right) + 2\mu \left((a^2 - 2a^2 \cos(-\varepsilon))^{1/2} \times \frac{na^2 \sin(-\varepsilon)}{(a^2 - 2a^2 \cos(-\varepsilon))^{1/2}} \right)$$

$$+ \sigma \left(a^2 - 2a^2 \cos(-\varepsilon) \right) = 0 \qquad (2.47)$$

$$\rho V \left(\frac{n^2 a \cos(-\varepsilon)}{(1 - 2\cos(-\varepsilon))^{1/2}} - \frac{n^2 a \sin^2(-\varepsilon)}{(1 - 2\cos(-\varepsilon))^{3/2}} \right) + 2\mu \left(na^2 \sin(-\varepsilon) \right) + \sigma \left(a^2 (1 - 2\cos(-\varepsilon)) \right) = 0$$

$$(2.48)$$

To linearize (2.48) we multiply through it by $\left(1 - 2\cos(-\varepsilon)\right)^{\frac{3}{2}}$ such that

$$\rho V \left(n^2 a \cos(-\varepsilon)(1 - 2\cos(-\varepsilon)) - n^2 a \sin^2(-\varepsilon) \right) + 2\mu \left(na^2 \sin(-\varepsilon)(1 - 2\cos(-\varepsilon))^{3/2} \right) + \sigma \left(a^2 (1 - 2\cos(-\varepsilon))^{5/2} \right) = 0$$

(2.49)

Note that $\cos(-\varepsilon) = \cos\varepsilon$ (even and symmetric function) and $\sin(-\varepsilon) = -\sin\varepsilon$ (odd and screw symmetric function), as a result (2.49) yields the following result.

$$\rho V \left(n^2 a \cos\varepsilon(1 - 2\cos\varepsilon) - n^2 a \sin^2\varepsilon \right) + 2\mu \left(-na^2 \sin\varepsilon(1 - 2\cos\varepsilon)^{3/2} \right) + \sigma \left(a^2 (1 - 2\cos\varepsilon)^{5/2} \right) = 0$$
(2.50)

Later we are going to utilize two types of approximation to reduce the fractional index in (2.50) since we are dealing with micro-vascular blood vessels.

2.2 Calculation of the elasticity of the blood medium (σ)

We know that the elasticity of the human aorta is about $\mu = 10 \times 10^5 \, dyne/cm^2 = 1 \times 10^5 \, N/m^2$ or more explicitly written as $= 1 \times 10^5 \, kgm^{-1}s^{-2}$ since (1N=105dyne, 104 cm2 = 1m2). Given that the dynamic viscosity of the human blood is about $\eta = 0.004 kgm^{-1}s^{-1}$ and also the approximate angular frequency of the human heart is $f = 1.2\,s^{-1}$. With the provision of these parameters we can calculate the elasticity of the blood medium σ from the equation

$$\sigma = \frac{\eta^2 f^2}{\mu} = \frac{(0.004 kg\, m^{-1}s^{-1})^2 \times (1.2 s^{-1})^2}{1 \times 10^5 kgm^{-1}s^{-2}} = 2.304 \times 10^{-10} \, kgm^{-1}s^{-2}$$

(2.51)

2.3 Calculation of the phase angle (ε), angular frequency (n) and the amplitude (a) of the 'host wave'.

We have said that two types of approximation shall be utilized in order to linearize (2.50). These are the 'third and the fourth world approximation'. These approximations are the differential minimization of the resulting binomial expansion of a given variable function. These approximations have the advantage of converging results easily and also producing expected minimum value of results. Now the 'third world approximation' states that

$$\left(1 + \xi f(\phi)\right)^{\pm n} = \frac{d}{d\phi}\left(1 + n\xi f(\phi) + \frac{n(n-1)}{2!}\left(\xi f(\phi)\right)^2 + \frac{n(n-1)(n-2)}{3!}\left(\xi f(\phi)\right)^3 + \ldots\right) - n\frac{d}{d\phi}\left(\xi f(\phi)\right)$$

(2.52)

While the 'fourth world approximation' states that

$$\left(1 + \xi f(\phi)\right)^{\pm n} = \frac{d}{d\phi}\left(1 + n\xi f(\phi) + \frac{n(n-1)}{2!}\left(\xi f(\phi)\right)^2 + \frac{n(n-1)(n-2)}{3!}\left(\xi f(\phi)\right)^3 + \ldots\right)$$

$$- n\frac{d}{d\phi}\left(\xi f(\phi)\right) - \frac{n(n-1)}{2!}\frac{d}{d\phi}\left(\xi f(\phi)\right)^2 - \cdots$$

(2.53)

Here ϕ is any variable function and ξ is some scalar number. The 'fourth world approximation' enhances minimum functional value since the amplitude of the constituted carrier wave would have to go through the smallest blood vessel – the capillaries. Hence by using the 'third world approximation' in (2.50) we get

$$(1 - 2\cos\varepsilon)^{3/2} = (1 + (-2\cos\varepsilon))^{3/2} = (0 + 3\sin\varepsilon - 3\cos\varepsilon\sin\varepsilon + \ldots) - 3\sin\varepsilon = -3\cos\varepsilon\sin\varepsilon \quad (2.54)$$

$$(1 - 2\cos\varepsilon)^{5/2} = (1 + (-2\cos\varepsilon))^{5/2} = (0 + 5\sin\varepsilon - 15\cos\varepsilon\sin\varepsilon + \ldots) - 5\sin\varepsilon = -15\cos\varepsilon\sin\varepsilon \quad (2.55)$$

Also by following the same algebraic subroutine in (2.50), the 'fourth world approximation' yields the below result.

$$(1 - 2\cos\varepsilon)^{3/2} = (1 + (-2\cos\varepsilon))^{3/2} = -\frac{3}{2}\cos^2\varepsilon\sin\varepsilon \quad (2.56)$$

$$(1 - 2\cos\varepsilon)^{5/2} = (1 + (-2\cos\varepsilon))^{5/2} = \frac{15}{2}\cos^2\varepsilon\sin\varepsilon \quad (2.57)$$

When we substitute (2.54) and (2.55) into (2.50) and after possible simplifications we realize that

$$\rho V n^2 a \left(\cos\varepsilon - 2\cos^2\varepsilon - \sin^2\varepsilon\right) + a^2 \left(6\mu n\cos\varepsilon\sin^2\varepsilon - 15\sigma\cos\varepsilon\sin\varepsilon\right) = 0 \quad (2.58)$$

We can now equate the coefficient of the terms in the parentheses to zero so that we get two separate results as

$$\left(\cos\varepsilon - 2\cos^2\varepsilon - \sin^2\varepsilon\right) = 0 \quad (2.59)$$

$$n = \frac{15\,\sigma\cos\varepsilon\sin\varepsilon}{6\mu\cos\varepsilon\sin^2\varepsilon} = \frac{15\,\sigma}{6\mu\sin\varepsilon} \quad (2.60)$$

Also by a similar substitution of (2.56) and (2.57) into (2.50) we get after a careful simplification

$$2\rho V n^2 a \left(\cos\varepsilon - 2\cos^2\varepsilon - \sin^2\varepsilon\right) + a^2 \left(6\mu n\cos^2\varepsilon\sin^2\varepsilon + 15\sigma\cos^2\varepsilon\sin\varepsilon\right) = 0 \quad (2.61)$$

$$a = \frac{-\left(\cos\varepsilon - 2\cos^2\varepsilon - \sin^2\varepsilon\right)2\rho V n^2}{6\mu n\sin^2\varepsilon\cos^2\varepsilon + 15\sigma\cos^2\varepsilon\sin\varepsilon} \quad (2.62)$$

Let us now solve for the critical value of the phase angle of the 'host wave' by using the relation $\cos\varepsilon = 1 - \frac{\varepsilon^2}{2}$ and $\sin\varepsilon = \varepsilon$ in (2.58), so that

$$\varepsilon^4 - \varepsilon^2 + 2 = 0 \quad (2.63)$$

Upon solving for ε in (2.63) we get four roots as its possible solutions given by

$$\varepsilon_1 = -0.9783 + 0.6761i\; ;\; \varepsilon_2 = -0.9783 - 0.6761i\; ;\; \varepsilon_3 = 0.9783 + 0.6761i\; ;\; \text{and}\; \varepsilon_4 = 0.9783 - 0.6761i \quad (2.64)$$

Thus a more realistic complex value of ε is $\varepsilon_3 = 0.9783 + 0.6761i$ and by converting the result from the complex value to a value in degree we get

$$\tan\varepsilon = \frac{0.6761}{0.9783} = 0.6911 \;\Rightarrow\; \varepsilon = \tan^{-1}(0.6911) = 0.6109\;rad.\;(35^0) \quad (2.65)$$

$$n = \frac{15 \times 2.304 \times 10^{-10}\,kgm^{-1}s^{-2}}{6 \times 0.004 \times \sin(0.6109)\,kgm^{-1}s^{-1}} = 2.51 \times 10^{-7}\;rad./s \quad (2.66)$$

$$a = \frac{-\left(\cos(0.6109) - 2\cos^2(0.6109) - \sin^2(0.6109)\right)2\rho V n^2}{6(2.51 \times 10^{-7})(0.004)\cos^2(0.6109)\sin^2(0.6109) + 15(2.304 \times 10^{-10})\cos^2(0.6109)\sin(0.6109)} \quad (2.67)$$

$$a = \left(3202389831\,kg^{-1}ms^2\right)2\rho V n^2 \quad (2.68)$$

Now for the human ascending aorta whose radius $r = 0.015m$ and length $l = 0.07m$, then the volume V is

$$V = \pi r^2 l = 3.142 \times (0.015)^2 \times 0.07 = 4.94865 \times 10^{-5} m^3 \quad (2.69)$$

$$a = \left(3202389831 kg^{-1} m s^2\right) \times 2 \times 1050 kgm^{-3} \times 4.94865 \times 10^{-5} m^3 \times (2.51 \times 10^{-7} rad.s^{-1})^2 = 2.1 \times 10^{-6} m \quad (2.70)$$

Also the amplitude of the carrier wave in the human capillary can be calculated from (2.68). The length of the human capillary is about 600 X10-6m and the diameter is about 10X10-6 m (radius r = 5X10-6 m), the approximate value of the amplitude is $a = 4.99 \times 10^{-12} m$. However, we are not going to use this value in our computation since we have assumed that the blood spends a negligible time in the capillary during the energy exchange process

2.4   Calculation of the wave number or spatial frequency (k)of the 'host wave'.

We have made the assumption that for the carrier wave to have a maximum value then the spatial oscillating phase must be equal to 1, as a result

$$\phi = \cos\left((\vec{k} - \vec{k}'\lambda).\vec{r} - (n - n'\lambda)t - |\partial E|\right) = 1 \quad (2.71)$$

$$\left((\vec{k} - \vec{k}'\lambda).\vec{r} - (n - n'\lambda)t - |\partial E|\right) = 0 \quad (2.72)$$

$$(\vec{k} - \vec{k}'\lambda) = (k - k'\lambda)_x i + (k - k'\lambda)_y j + (k - k'\lambda)_z k \quad (2.73)$$

$$\vec{r} = xi + yj + zk \quad (2.74)$$

If we assume that the motion is constant in the z-direction and the wave vector mode is also the same for both x and y plane in the cylindrical system then with the usual transformation from Cartesian to polar coordinate system ($x = r\cos\theta \ and \ y = r\sin\theta$), we get

$$\vec{r} = r\cos\theta \, i + r\sin\theta \, j \quad (2.75)$$

$$(\vec{k} - \vec{k}'\lambda) = (k - k'\lambda)_x i + (k - k'\lambda)_y j \quad (2.76)$$

where $\theta = \pi - (\varepsilon - \varepsilon'\lambda)$ is the variable angle between y_1 and y_2 , please see appendix for details.

$$(\vec{k} - \vec{k}'\lambda).\vec{r} =\underline{\ } (k - k'\lambda) \, r \, (\cos\theta + \sin\theta) \quad (2.77)$$

The constituted carrier wave can only have a maximum value provided the oscillating phase is equal to one, that is

$$\phi = \cos\left((k - k'\lambda) \ r \, (\cos\theta + \sin\theta) - (n - n'\lambda)t - |\partial E|\right) = 1 \quad (2.78)$$

$$\left((k - k'\lambda)r(\cos\theta + \sin\theta) - (n - n'\lambda)t - |\partial E|\right) = 0 \quad (2.79)$$

Using the boundary conditions that at time $t = 0$, $\lambda = 0$, $|\partial E| = \varepsilon = 0.6109 rad.$

$$\theta = \pi - (\varepsilon - \varepsilon'\lambda) = \pi - \varepsilon = 3.142 - 0.6109 = 2.5311 rad. \quad (2.80)$$

$$k \, r \left(\cos(2.5311) + \sin(2.5311)\right) - 0.6109 rad. = 0 \quad (2.81)$$

$$k = \frac{0.6109 rad.}{r \, (-0.2460945)} = \frac{0.6109 rad.}{0.015m \times (-0.2460945)} = 166 rad./m \quad (2.82)$$

Note that the radius of the human aorta r = 0.015m and we are working with the absolute value of k .

2.5   Calculation of the phase angle (ε'), angular frequency (n'), wave number (k') and amplitude (b)of the HIV/AIDS parameters

The gradual deterioration in the intrinsic parameters of the biological system of a HIV/ AIDS infected candidate would make us believe that after a sufficiently long period of time all the active constituents of the biological system of the host (Man) would have been completely eroded by the influence of the HIV, on the basis of this argument the below relation holds

$$
\left.\begin{array}{l}
a - b\lambda = 0 \ \Rightarrow 2.1 \times 10^{-6} = b\lambda \\[6pt]
n - n'\lambda = 0 \Rightarrow 2.51 \times 10^{-7} = n'\lambda \\[6pt]
\varepsilon - \varepsilon'\lambda = 0 \Rightarrow 0.6109 = \varepsilon'\lambda \\[6pt]
k - k'\lambda = 0 \Rightarrow 166 = k'\lambda
\end{array}\right\} \quad (2.83)
$$

Upon dividing the above relations in (2.83) with one another to eliminate λ, we get the following six relations.

$$
\left.\begin{array}{l}
8.3665n' = b \\[6pt]
3.43755 \times 10^{-6}\,\varepsilon' = b \\[6pt]
1.26506 \times 10^{-8}\,k' = b \\[6pt]
4.10869 \times 10^{-7}\,\varepsilon' = n' \\[6pt]
1.51204 \times 10^{-9}\,k' = n' \\[6pt]
3.68012 \times 10^{-3}\,k' = \varepsilon'
\end{array}\right\} \quad (2.84)
$$

Thus from all indications k' is related to ε' according to; $3.68 \times 10^{-3} k' = \varepsilon'$ or $3.68k' = 10^3 \varepsilon'$. Suppose we want to re-establish another relationship between k' and ε' then we can simply multiply through the sets of the relations in (2.84) by 3.68012×10^3. Consequently, once this is done then a more realistic and applicable relation for both quantities can be found from the 2nd and 3rd relations in (2.84) after equating them, that is, when $0.0126505985\varepsilon' = 0.000046555k'$. Hence from simple ratio we can generally establish that

$$
\varepsilon' = 0.0000466\,rad.; \quad k' = 0.0127\,rad./m; \quad n' = 1.91 \times 10^{-11}\,rad./s; \quad b = 1.60 \times 10^{-10}\,m \quad (2.85)
$$

Any of these estimated values of the HIV parameters shall produce a corresponding approximate value of the multiplicative factor $\lambda = 13070$ upon substituting them into (2.83). Hence the range of the multiplicative factor is $0 \le \lambda \le 13070$.

2.6 Determination of the attenuation constant (η)

Attenuation is a decay process. It brings about a gradual reduction and weakening in the initial strength of the basic intrinsic parameters of a given physical system. In this study, the parameters are the amplitude (a), phase angle (ε), angular frequency (n) and the spatial frequency (k). The dimension of the attenuation constant (η) is determined by the system under study. However, in this work, attenuation constant is the relative rate of fractional change σ (FC) in the basic parameters of the carrier wave function. There are 4 (four) attenuating parameters present in the carrier wave. Now, suppose a, n, ε, k represent the initial parameters of the 'host wave' that is present in the carrier wave and $a - b\lambda$, $n - n'\lambda$, $\varepsilon - \varepsilon'\lambda$, $k - k'\lambda$ represent the final parameters of the 'host wave' that survives after a given time. Then, the FC is

Fractional change,
$$
\sigma = \frac{1}{4} \times \left(\left(\frac{a - b\lambda}{a} \right) + \left(\frac{\varepsilon - \varepsilon'\lambda}{\varepsilon} \right) + \left(\frac{n - n'\lambda}{n} \right) + \left(\frac{k - k'\lambda}{k} \right) \right) \quad (2.86)
$$

$$
\eta = \frac{FC\big|_{\lambda=i} - FC\big|_{\lambda=i+1}}{unit\ time\,(s)} = \frac{\sigma_i - \sigma_{i+1}}{unit\ time\,(s)} \quad (2.87)
$$

The dimension is per second (s^{-1}). Thus (2.81) gives $\eta = 0.0000763\,s^{-1}$ for all values of $\lambda\,(i = 0, 1, 2, 3, \ldots, 13070)$.

2.7   Determination of the attenuation time (t)

We used the information provided in section 2.6, to compute the various times taken for the constitutive carrier wave to attenuate to zero. The maximum time the carrier wave lasted as a function of the raising multiplier λ is also calculated from the attenuation equation shown by (2.87). The reader should note that we have adopted a slowly varying regular interval for the multiplicative factor λ for our study. The slow varying interval we adopted will help to delineate clearly the physical parameter space accessible to our model. However, it is clear from the calculation that different attenuating fractional changes contained in the carrier wave function are approximately equal to one another. We can now apply the attenuation time equation given below.

$$\sigma = e^{-(\alpha \eta t)/\lambda} \qquad (2.88)$$

$$t = -\left(\frac{\lambda}{\alpha \eta}\right)\ln \sigma \qquad (2.89)$$

In this case, α is the HIV index factor andfor a non-local biological diseasesuch as the HIV $\alpha = 3$. However, the factor α is different for any other human biological diseases that are not localized. The equation is statistical and not a deterministic law, it gives the expected intrinsic parameters of the 'host wave' that survives after time t. Generally, we used the table scientific calculator and Microsoft Excel to compute our results. Also the GNUPLOT 4.6 version was used to plot the corresponding graphs.

III.   PRESENTATION OF RESULTS
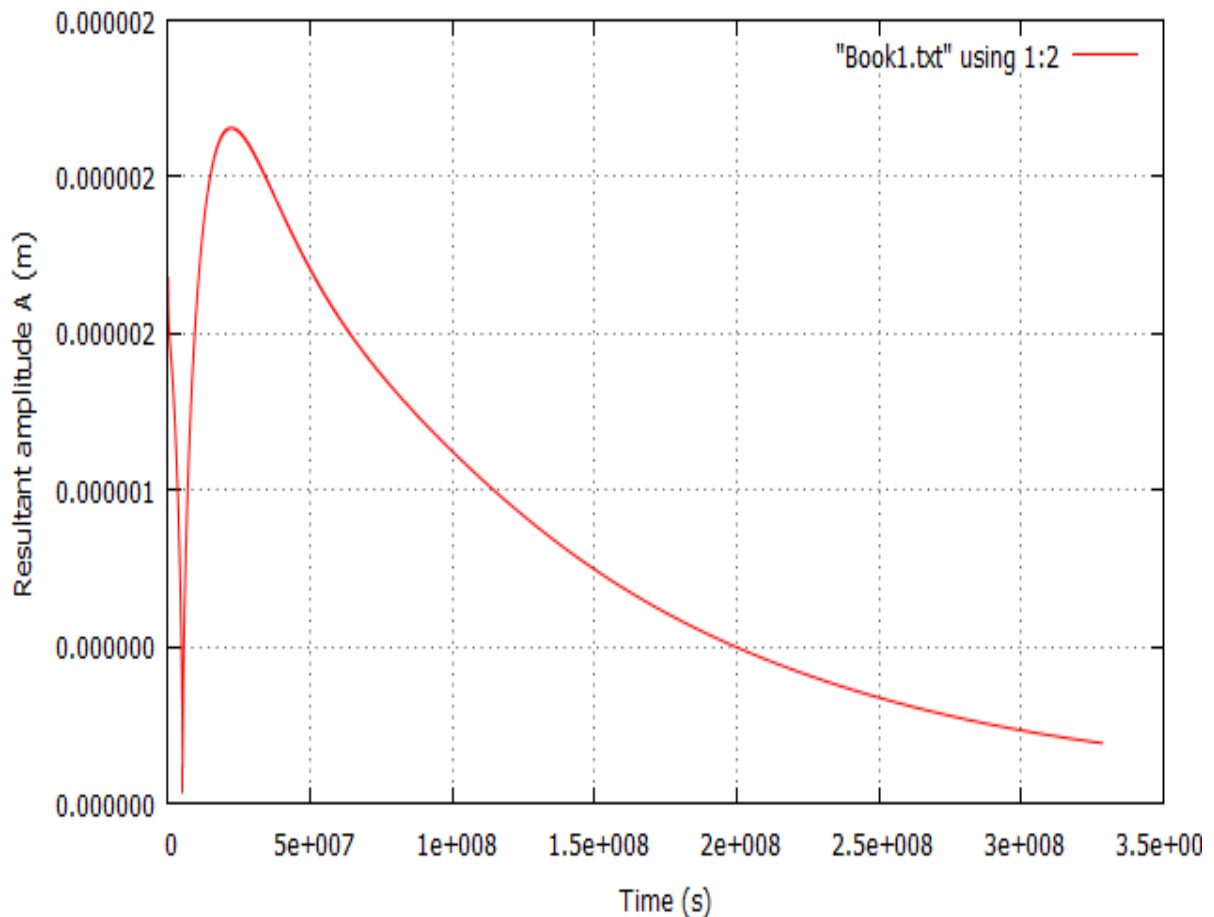
Fig. 1: Represents the graph of the resultant amplitude A of the carrier wave y as a function of time t.
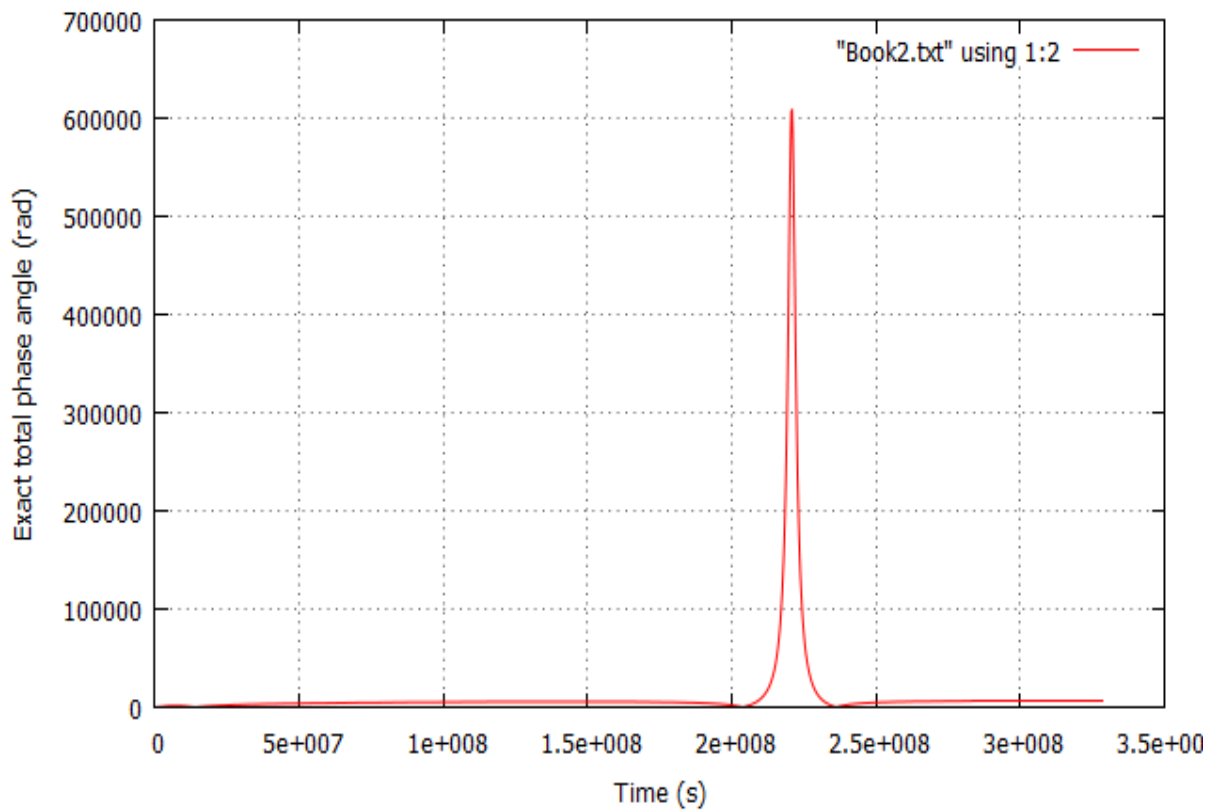
Fig. 2: Represents the graph of the exact total phase angle $\left|\partial E\right|$ of the carrier wave ^{y} as a function of time ^{t} .
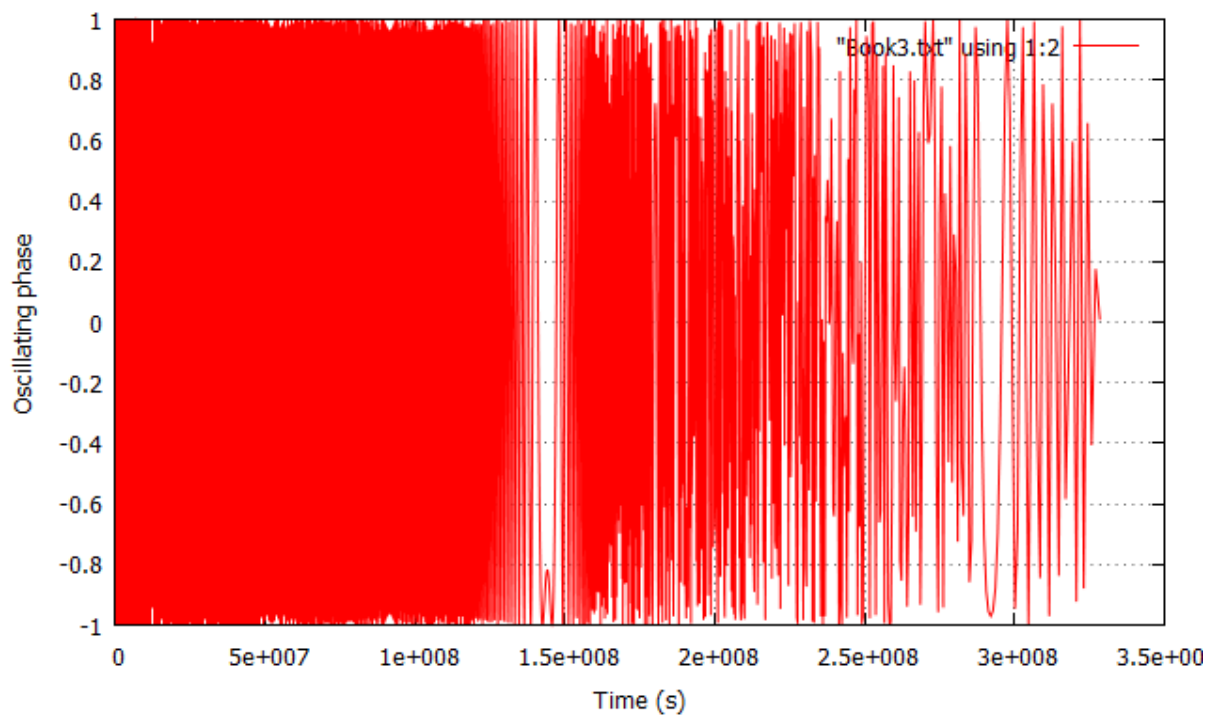
Fig. 3: Represents the graph of the spatial oscillating phase $^\phi$ of the carrier wave ^{y} as a function oftime ^{t} .
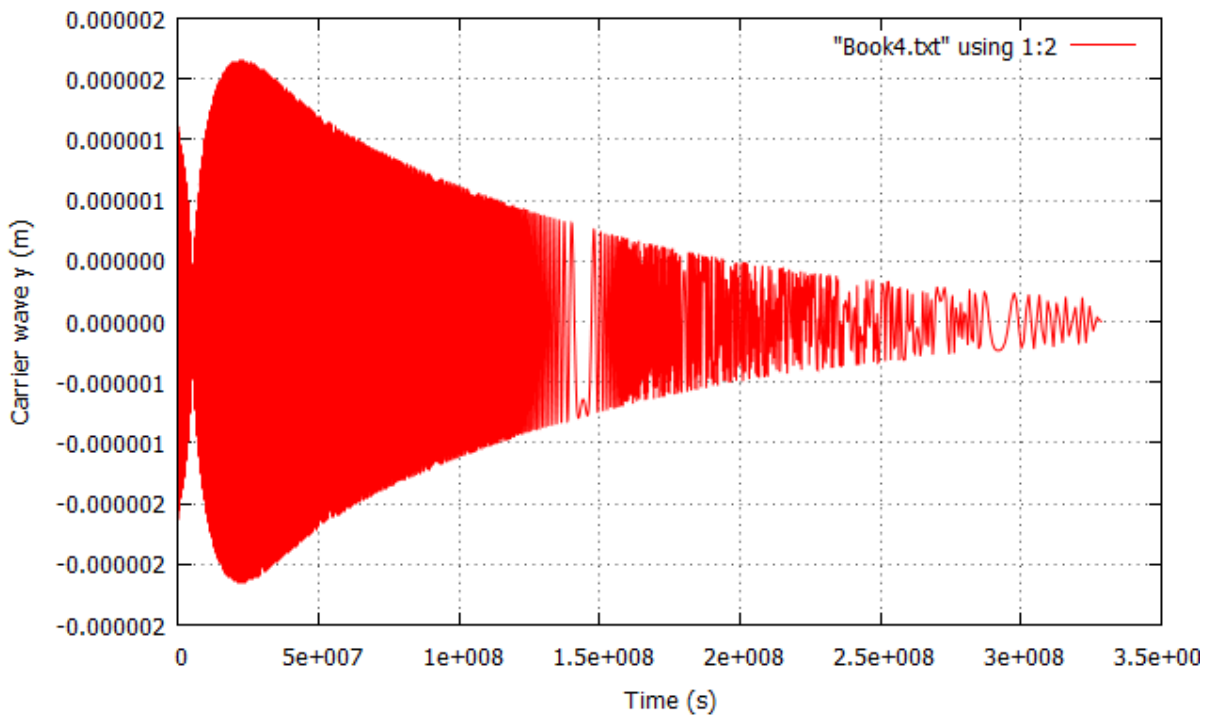
Fig. 4: Represents the graph of the carrier wave y as a function of time t.

IV. DISSCUSION OF RESULTS

It is clear from fig. 1, that within the interval of the multiplicative factor $0 \le \lambda \le 3627$ the absolute value of the resultant amplitude of the carrier wave first decreases consistently to a minimum value of 6.6564x10-9m when the time is 5133294s (2 months).When the multiplicative factor $3627 \le \lambda \le 6874$, the amplitude then rises with real values to a maximum value of 2.15862 10-6m at time t = 22837600s (8.8 months). This characteristic is also reflected in fig.4.Thus within the range of the raising multiplier $0 \le \lambda \le 6874$ the HIV would have lasted for about 8 months in the human system. After this time, the amplitude steadily decreases exponentially to zero after a prolonged time of328479340s (10 years and 4 months).There is need for us to emphasize here that the initial value of the amplitude of the 'host wave' which we calculated to be 2.1X10-6 m is the static value. However, in practice due to the non-stationary oscillating behavior of the amplitude it turns out that the absolute value of the 'host wave' amplitude is 1.677X10-6 m.

From our calculation that within the interval of the multiplicative factor $0 \le \lambda \le 3627$ and the time $0 \le t \le 5133294s$, the amplitude of the carrier wave initially experiences a steady imaginary decrease in value from $i\,1.6777 \times 10^{-6} m$ to $i\,6.6564 \times 10^{-9} m$. This behaviour is also reflected in the Fig.4. However, the reader

should note that it is the absolute values of the amplitude that we used in the graphical presentation. From our calculation therefore, it takes about 2 to 8 months for the HIV parasite to incubate before its absolute manifestation is felt. This according to the literature of clinical disease is the window period. The window period therefore, is the difference in time between the end of the imaginary value of the amplitude and the maximum positive value that the amplitude attained. In other words, the HIV 'parasitic wave' does not take immediate absolute effect on the human system when it is contacted. Within this interval of time, there is constant agitation by the intrinsic parameters of the 'host wave' to resist, thereby suppressing the destructive influence of the interfering HIV 'parasitic wave'. During this period, although unnoticeable as it may, but much imaginary harm would have been done to the constituent parameters of the biological system of the host (Man). Consequently, the amplitude of the carrier wave is actually made up of the imaginary and real part, $A = A_1 + iA_2$. This shows that the motion is actually two-dimensional (2D). Thus A_1 and A_2 are the components of the amplitude in x and y - directions, and A is tangential to the path of the moving amplitude in the carrier wave.

The graph of the exact total phase angle $|\partial E|$ ofthe carrier wave as a function of time as shown in fig. 2, almost show undefined behaviour everywhere except at time t = 220862000 s (7 years) and this occurs when $|\partial E|$ =604442 rad. The predominant behaviour of the carrier wave at this particular time is referred to as the wave packet. The wave packet is a localized pulse that is composed of both the 'host wave' and the 'parasitic wave' that cancel each other everywhere else except at this time during the interference. Since each component of the wave packet has different phase velocity in the medium, the modulation propagation number $(k - k'\lambda)$ of the components of the carrier wave changes in the medium and consequently the group velocity changes. This results to a change in the width of the wave packet.

It is clear from both figures that the events associated with the spatial oscillating phase and the carrier wave is similar. Initially, the spectra of fig. 3 and 4 are both blurred up to 4.5 years(144112000s) followed by a gradual depletion of the wave form. The blurred nature of the spectra is an indication of the resistance of the intrinsic parameters of the 'host wave' to the destructive tendency of the interfering 'parasitic wave'. While the subsequent depleting behaviour involves a steady decay process, resulting to a gradual reduction and weakening in the initial strength of the intrinsic parameters of the host biological system. The spatial oscillating phase oscillates between the maximum and minimum values of +1 and -1. In phasor language, for positive spatial oscillating phase y_1 leads y_2 and ε leads ε' (or ε' lags ε)in the CCW, while for negative value, y_2 leads y_1 and ε' leads ε (or ε lags ε').

As we can observe, the spectrum of fig. 3 and 4 show an exemplary behaviour or a sharp transition from one system phase to another around 1.44112×10^{8} s (4.5 years).After a prolonged time, the spectra of figs. 3 and 4, again experiences a more pronounced depletion after 2.8x108s (8.8 years) and the constituents of the carrier wave becomes monochromatic in nature, that means a predominance of the 'host wave' after this time. In this case, the group velocity of the carrier wave becomes equal to the phase velocity.Consequently, it is obvious from both figures that the infected biological system could possibly degenerate from HIV infection to a phenomenon of AIDS between 4.5 years or 8.8 years in the absence of specific treatment.

The attenuating behaviour of the carrier wave in the interval $0 \leq t \leq 5133294\,s$ is a consequence of the fact that the carrier wave do not steadily go to zero, rather it fluctuates. The fluctuation is due to the constructive and destructive interference of both the 'host wave' and the 'parasitic wave' contained in the carrier wave. In the regions where the amplitude of the carrier wave is greater than either of the amplitude of the individual wave, we have constructive interference that means the path difference is $(\varepsilon + \varepsilon'\lambda)$, otherwise, it is destructive in which case the path difference is $(\varepsilon - \varepsilon'\lambda)$. If $n \approx n'$, then the average angular frequency say $(n + n')/2$ will be much more greater than the modulation angular frequency say $(n - n')/2$ and once this is achieved, then we will have a slowly varying carrier wave with a rapidly oscillating phase.

The spectrum of the displacement vector of the carrier wave as a function of time shown in fig. 4 is similar to that of fig. 1, since the carrier wave is the product of the resultant amplitude A and the oscillating phase ϕ. The exception to this similarity is that the carrier wave oscillates non-consistently with quadratic dispersion before it finally comes to rest.The carrier wave first decreases from the initial absolute value of 1.6777x10-6m to 2.05156x10-9m for a period of time t = 5133294s (2 months). Then the carrier wave oscillates to a maximum and minimum value of 2.14358x10-6m and -2.17386x10-6m respectivelyat a time t=22473700s (8.6 months).The displacement vector of the carrier wave decreases from the initial absolute value of 1.6777x10-6 m to 2.05156x10-9m. Hence the phenomenon called acquired immunodeficiency syndrome AIDS, actually occurs in the interval when the multiplicative factor $\lambda \geq 12803$and the time t $\geq$ 2.8x108s (8.8 years). Thus in this region, the system of the infected candidate can no longer recover from the HIV infection irrespective of any treatment administered. The resistance to treatment which means lack of immunity is as a result of the fact that all the active constituents of the 'host wave' would have been completely attenuated by the influence of the interfering 'parasitic wave'which is now the predominant constituent of the carrier wave.

Our calculation shows that in the absence of specific treatment, the HIV infection degenerates to AIDS after either4 years or 8 years and that is when the multiplicative factor $12803 \leq \lambda \leq 13070$. This period involves a steady decay process which results to a gradual reduction and weakening in the initial strength of the

intrinsic parameters of the host biological system. In this casethe displacement of the carrier wave which describes the coexistence of the biological system of Man and the HIV ceases to exist – the phenomenon called deatharound $328479340s$ (10 years)and the multiplicative factor λ would have attained the critical value of 13070.

V. CONCLUSION

Initially, the carrier wave spectrum shows a blurred characteristics followed by a gradual depletion of the wave form. The blurred nature of the resulting spectra is an indication of the resistance of the intrinsic parameters of the 'host wave' to the destructive tendency of the interfering 'parasitic wave'. While the subsequent depleting behaviour means a predominance of the 'parasitic wave'. After this time, a steady decay process resulting to a gradual reduction and weakening in the initial strength of the intrinsic parameters of the host biological system becomes prominent. The constituents of the carrier wave becomes monochromatic in nature since each component of the wave packet has different phase velocity in the medium, the modulation propagation number $(k - k'\lambda)$ of the components of the carrier wave changes in the medium and consequently the group velocity changes. This results to a change in the width of the wave packet. It seems from the results, that the actual dynamic components of the HIV responsible for their destructive tendency are $b\lambda$, $n'\lambda$, $\varepsilon'\lambda$ and $k'\lambda$. This study revealed that in the absence of specific treatment, people infected with HIV develop AIDS within 8 years and the average survival time after infection with HIV is found to be 8 to 10 years. Also the cessation of the carrier wave which describes the coexistence of the system of Man and the HIV is not instantaneous but gradual. Consequently, the life span of any biologically active system is determined by the resistance of its intrinsic parameters to the destructive influence of any internal or external factor. Finally, the significance of the HIV pandemic has been scientifically exaggerated, since the search strategy for the cure may lie outside the propounded complex scientific ideas and approach. Limitations in this work would include the use of some physical data such as the dynamic viscosity of blood whose values may not be exact based on small variations which differs from one individual to another. Secondly, the study relied on measures of sensitive physical parameters such as the radius and length of the human aorta, which may also differ slightly between individuals. Thus the findings presented here are approximate and may not be generalizable to the greater sense.APPENDIX: Vector representation of the superposition of the 'host wave' and the 'parasitic wave'. The amplitude of the CCW both waves are not constant with time but they oscillate at a given frequency.
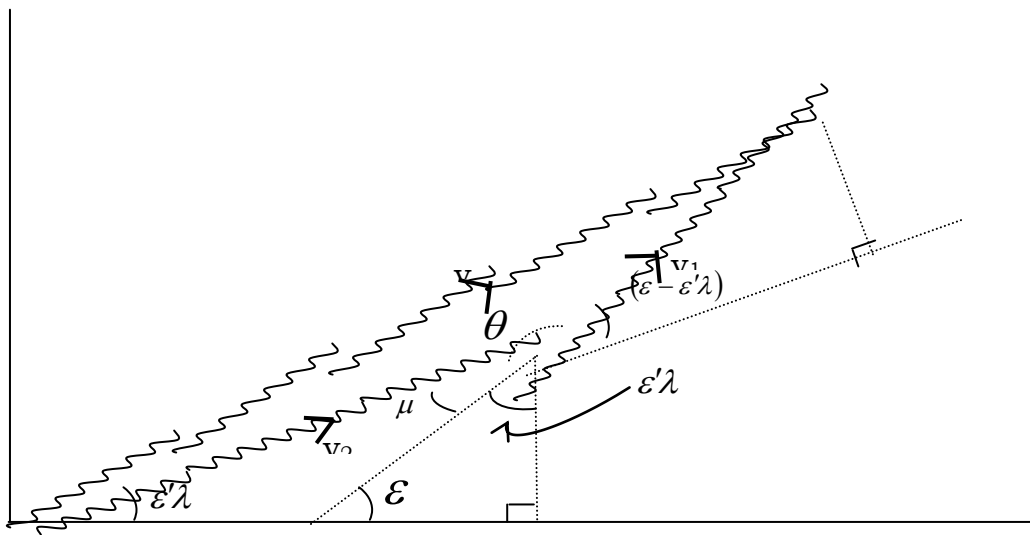
Fig. A 1:Represents the human 'host wave' y_1 and the HIV 'parasitic wave' y_2 after the interference. The superposition of both waves y_1 and y_2 is represented by the carrier wave displacement y. It is clear that from the geometry of the figure: $\mu + \varepsilon'\lambda + 180^0 - \varepsilon = 180^0$; $\mu = \varepsilon - \varepsilon'\lambda$; $\theta = 180^0 - (\varepsilon - \varepsilon'\lambda)$; and $\theta = \pi - (\varepsilon - \varepsilon'\lambda)$.

REFERENCES

[1]   Sepkowitz, K. A. (2001) 'AIDS-the first 20 years'
[2]   N. Engl. J. Med. 344 (23) : 1764 – 72.
[3]   Mandel, Bennet, and Dolan (2010)
[4]   HIV/AIDS – Wikipendia, Chapter 118.
[5]   UNAIDS, WHO (December 2007) '2007 AIDS Epidemic Update' (PDF)
[6]   The Aorta – Grays anatomy of the human body.
[7]   http://education.yahoo.com
[8]   Mette S. Olufsen, Charles S. Peskin, Won Yong Kim, Erik M. Pedersen, Ali Nadim and Jesper Larsen (2000). Numerical simulation and Experimental validation of Blood flow in Arteries with structured-Tree outflow conditions.
[9]   Annals of Biomedical Engineering, Vol. 28, pp. 1281 – 1299.
[10]   Stefan Anitei (softpedia) (December 2007). '7 Things about the heart and blood vessels
[11]   www.google.com.ng
[12]   Cutnel, John and Johnson, Kennet
[13]   Physics, 4th Edition, Wiley, 1998, p 308.
[14]   Alexander, R. W., Schlant, R. C., and Fuster, V. 'Hurst's the Heart, Arteries and Veins.
[15]   9th Edition McGraw-Hill, New York, NY. 1998.
[16]   Glenn Elert (2008). 'Viscosity'
[17]   http://hypertextbook.com/physics/matter/viscosity
[18]   Tzu-Ching Shih et al. (2012) 'Numerical analysis of Coupled effects of pulsatile blood flow and thermal relaxation time during thermal therapy.
[19]   International Journal of Heat and mass transfer, 55: 3763 – 3773
[20]   DavidHalliday, Robert Resnick and Jearl Walker
[21]   Fundamentals of Physics, 6th Edition, John Weley and Sons, Inc. New york 2001 : p378.
[22]   Valenta J.(1993). In Biomechanics, clinical aspects of biomedicine, 2, Elsevier, Amsterdam.

A Green Capacitated Vehicle Routing Problem with Fuel Consumption Optimization Model

1,İlker Küçükoğlu , 2,Seval Ene, 3,Aslı Aksoy[*] , 4,Nursel Öztürk

1,2,34,Uludag University Faculty of Engineering and Architecture   Industrial Engineering Department
Gorukle Campus, 16059, Bursa, Turkey
[*]Corresponding author asliaksoy@uludag.edu.tr

ABSTRACT:

In recent years, since environmental issues and regulations impact strategic and operational decisions of companies, green logistics has a critical and gaining value for researchers and companies. In this context green capacitated vehicle routing problem (G-CVRP) under minimizing fuel consumption, which is rarely encountered in real life systems, is studied in this paper. A mixed integer linear programming model is proposed for solving G-CVRP. In the G-CVRP optimization model, fuel consumption is computed considering the vehicle technical specifications, vehicle load and the distance. Fuel consumption equation is integrated to the model through a regression equation *proportional to the distance and vehicle load. G-CVRP optimization model is validated by various instances with different number of customers. Achieved results show that G-CVRP model provides important reductions in fuel consumption.*

KEYWORDS: *Fuel consumption model, green vehicle routing, mixed integer linear programming.*

I.   INTRODUCTION

In recent years, with the growing environmental concerns, industrial organizations have to take into account environmental factors in order to enhance the competitive aspect. In this global world, logistics stays at the center of the modern transportation systems. As the competition increases, the technology rapidly changes, product life-cycle becomes shorter and the expectation of the customer increases. In high competitive environment, industrial organizations have to investigate logistics strategies and adopt green supply chain applications to consider accurate use of natural resources as a social responsibility. The importance of environmental issues is continuously translated into regulations, which potentially have a tangible impact on supply chain management. As a consequence, there has been an increasing amount of research on the intersection between logistics and environmental factors [1].

The scope of this study is, developing a mathematical model which minimizes the fuel consumption for green capacitated vehicle routing problem (G-CVRP). The study has two dependent components. The first component includes a graphical user interface to calculate fuel consumption for a given load and distance. The second component includes a mixed integer linear optimization model, to determine the route that minimizes fuel consumption. The optimization model works with the data obtained from fuel consumption calculation model. The remainder of this paper is organized as follows. Section 2 presents a literature review. Section 3 explains the fuel consumption calculation. Mixed integer linear programming model is presented in Section 4. Application examples and results are provided in Section 5. Finally, conclusions are presented in Section 6.

II.   LITERATURE REVIEW

Laporte described the vehicle routing problem (VRP) as a problem of designing optimal delivery or collection routes from one or several depots to a number of geographically scattered cities or customers, subject to side constraints [2]. VRP is a classic combinatorial optimization problem involved in many applications. The VRP plays a vital role in distribution and logistics. Since its introduction by Dantzig and Ramser, VRP has been extensively studied. By considering additional requirements and various constraints on route construction, different VRPs have been formulated, such as, pick-up and delivery VRP, capacitated VRP, multiple depots VRP, VRP with time windows. Although there are different forms of VRPs, most of them minimize the cost by

minimizing the total distance without considering the fuel consumption rate. In fact, statistics show that fuel cost is a significant part of total transportation cost [3].

There are many studies in the literature concerning with vehicle routing problem (VRP) [2, 4, 5, 6, 7 etc.]. In recent years, the interest in practical applications and literature about green logistics has increased; but, according to our knowledge, not much research has been conducted on the VRP under minimizing fuel consumption and emissions. Apaydın and Gönüllü studied the VRP model on reducing the CO_2 emissions vehicles used for waste collection operations in a city [8]. A shortest path model was used in order to optimize solid waste collection. Kuo, proposed a model for calculating total fuel consumption for the time-dependent vehicle routing problem (TDVRP) where speed and travel times are assumed to depend on the time of travel when planning vehicle routing. Simulated annealing (SA) algorithm is proposed for finding the vehicle routing with the lowest total fuel consumption [9]. Bektaş and Laporte described an approach to reduce energy requirements of vehicle routing based on a comprehensive emissions model that takes into account load and speed. The authors present a comprehensive formulation of the problem and solve moderately sized instances [10]. Suzuki, presented a model indicates that a significant saving in fuel consumption and CO_2 emissions may be realized by delivering heavy items in the early segments of a tour while delivering light items in the later segments, so that the distance a vehicle travels with heavy pay-loads can be minimized [11]. Wygonik and Goodchild developed a VRP with time windows to minimize emissions for urban pick-up and delivery system [12]. Xiao et al. proposed a mathematical model which minimizes fuel consumption of the vehicles. They developed a simulated annealing based algorithm to solve the model [3]. Erdogan and Hooks, considered fuel consumption of the vehicles and seek a set of vehicle tours that minimize total distance traveled to serve a set of customers while incorporating stops at alternative fueling stations in route plans [13].

III. FUEL CONSUMPTION CALCULATION

In this research, the proposed model aims to calculate total fuel consumption. Apaydın and Gönüllü used fixed values for fuel consumption [8]. Kuo; calculated fuel consumption with respect to vehicle load and vehicle speed [9]. Jabali et al. modeled fuel consumption as a function of speed [1]. Bektaş and Laporte calculated fuel consumption based on average speed and vehicle load [10]. Suzuki; considered vehicle load, average speed based on the road gradient and average amount of fuel consumed per hour while a vehicle is waiting at customer sites [11]. Ubeda considered vehicle load [14]. Xiao et al. formulated a linear function dependent on vehicle load for fuel consumption [3]. Huang et al. modeled linear regression for fuel consumption as a function of vehicle load and distance [7]. Demir et al. [15, 16] and Liu and Helfand used mathematical equation to calculate fuel consumption [17]. Table 1 indicates the summary of selected literature regarding criteria for fuel consumption and calculation methods.

Table 1. The summary of selected literature regarding criteria for fuel consumption and calculation methods

Reference	Criteria						Calculation Method		
	Fixed	Speed	Load	Road Gradient	Acceleration Rate		Regression	Table	Equation
Apaydın, Gönüllü [8]	+							+	
Jabali et al. [1]		+							+
Kuo [9]		+	+						+
Suzuki [11]		+	+	+					+
Ubeda et al. [14]			+					+	
Xiao et al. [3]			+				+		
Bektaş, Laporte[10]		+	+						+
Huang et al. [7]			+				+		
Demir et al. [15,16]		+	+		+				+
Liu, Helfand [17]	+								+

According to our knowledge, the previous studies ignored the acceleration rate and/or did not directly consider the vehicle load while calculating the fuel consumption. In this study, the fuel consumption is calculated considering the vehicle technical specifications, vehicle load and the route distance. This way of calculation yields more realistic approach. The vehicle needs energy to move with constant speed or acceleration in real time. This energy is equal to resistance forces. The resistance force includes rolling resistance (F_{Ro}) aerodynamic resistance (F_{Ae}), grade resistance (F_G) and acceleration resistance (F_{Acc}) [18]. The total force (F_T) at the wheels for a given acceleration and grade is

$$F_T = F_{Ro} + F_{Ae} + F_{Acc}$$

(1)

Assume that, P_T is the total power at the wheels and calculated as

$$P_T = F_T \, v \tag{2}$$

where v is the speed. To convert the total force to the fuel consumption, consumption value (b) multiplicities with the total power (P_T) as:

$$consumption = b \, P_T \tag{3}$$

To determine the time - speed graphic of a vehicle from node i to node j, it is assumed that, the vehicle accelerates from 0 to t_1, the speed is constant at speed limit (v_{max}) from t_1 to t_2, and decelerates from t_2 to t_3. Figure 1 shows the time - speed graphic which has three discrete areas: A_1 represents the area for the acceleration time, A_2 represents the area for the time with the constant speed and A_3 represents the area for the deceleration time. In this study, it is accepted that the acceleration rate of the vehicle is equal to the deceleration rate of the vehicle ($a_{acc} = a_{dec}$), thereby, the speed of the vehicle in real-time can be determined.
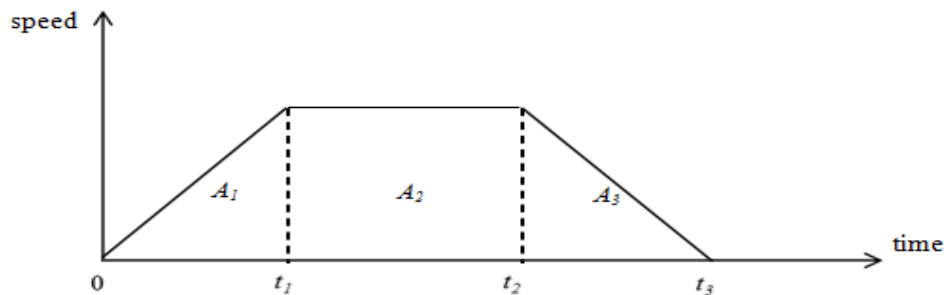
Figure 1. The speed vs. time graphic to calculate fuel consumption

In order to calculate the total fuel consumption of a vehicle in a specified distance, total power and consumption formulations are developed as a function of time with the following equations:

$$v(t) = \begin{cases} a_{acc} \, t, & if \ 0 \leq t < t_1 \\[2mm] v_{max}, & if \ t_1 \leq t < t_2 \\[2mm] v_{max} - a_{dec} \, (t - t_2), & if \ t_2 \leq t < t_3 \\[2mm] 0, & otherwise \end{cases} \tag{4}$$

$$F_T(t) = F_{Ro}(t) + F_{Ae}(t) + F_{Acc}(t) \tag{5}$$

$$F_T(t) = (fmg) + (c_d \, A \, \rho \, \frac{v^2(t)}{2})(\lambda \, ma) \tag{6}$$

$$P_T(t) = F_T(t) v(t) \tag{7}$$

$$Consumption = b P_T(t) \tag{8}$$

$$Total \ consumption = \int_0^{t_3} b P_T(t) \, d_t \tag{9}$$

where f is the coefficient of rolling resistance, m is the vehicle weight (empty plus carried load) and g is the gravitational constant, c_d is the aerodynamic coefficient, A is the frontal surface area, ρ is the air density and v is the speed, λ is the transmission variable ($\lambda = 1.04 + 0.0025 * i^2$, i is the overall gear ratio), a is the acceleration rate, α is the coefficient related to gradient of the road.

In this study, because of the complexity of the fuel consumption calculation (in equation 9), the parameters, except load and distance are assumed to be constant similar to Huang et al. [7]. So, the fuel consumption equation is proportional to the distance and vehicle load and regression analysis is achieved by using Minitab to obtain a mathematical equation between load, distance and fuel consumption, as shown:

$$Fuel \ consumption = a_1 \ distance + a_2 \ vehicle \ load + b \tag{10}$$

The data set used to achieve regression analysis, includes 2020 examples for different distances and vehicle loads in the range of 1300 – 1400 kg for vehicle weight and 0 – 100 km for distance. After the analysis, the calculated value of a_1, a_2 and b are; 0.0589, 0.000991, -1.5, respectively. The regression equation analysis is also performed in Minitab. The *R-square* value of the proposed regression model is 95.7%, indicating that the model fits the data extremely well.

IV. LINEAR INTEGER PROGRAMMING FORMULATION

The capacitated vehicle routing problem (CVRP) is an interesting combinatorial optimization task, which occurs frequently in real-world applications [19]. The CVRP may be described as the following graph problem. Let $G=(V,A)$ be a complete directed graph, where $V=\{0,1,\ldots,n\}$ is the vertex set and A is the arc set. Vertices $i=1,\ldots,n$ correspond to the customers, each with a known non-negative demand d_i and vertex 0 corresponds to the depot. A nonnegative cost c_{ij} is assigned to each arc $(i,j) \in A$, and it represents the cost of travelling from vertex i to vertex j. A set of K identical vehicles, each with the capacity C, is available at the depot. The CVRP consists of finding a collection of elementary cycles less than K in G with minimum total cost, such that:

- each cycle visits the depot vertex 0,
- each customer vertex $i=1,\ldots,n$ is visited by exactly one cycle,
- the sum of the demand d_i of the vertices visited by a cycle does not exceed the vehicle capacity C.

In this paper, a mixed-integer linear integer programming model is considered to construct routes for a set of vehicles to meet the demand of all customers, k homogenous vehicles depart from and return to the depot node after serving all the customers, each customer is visited at once and vehicles cannot carry load more than capacity. The overall objective is to minimize the travelling distance and fuel consumption. The green-capacitated vehicle routing problem (G-CVRP) optimization model, proposed in this research, can be developed as follows:

Indices

$i, j = 0,1,\ldots N$ Set of customers

$k = 1,\ldots K$ Set of vehicles

Parameters

c_{ij} Traveling distance between customer i and customer j

d_i Demand of customer i

Q Capacity of vehicles

a_1 Coefficient of regression equation

a_2 Coefficient of regression equation

b Coefficient of regression equation
M Big number
Min_load   Unladen weight of the vehicle
Max_load   *Min_load+Q*

Decision Variables

$x_{ijk} = \begin{cases} 1 & \text{if vehicle } k \text{ travels from customer } i \text{ to customer } j \quad (i \neq j) \\ 0 & \text{otherwise} \end{cases}$

$y_i = $ load of vehicle at customer i

Model

$$Min\ z = \sum_{i=0}^{N} \sum_{j=0}^{N} \sum_{k=1}^{K} a_1 c_{ij} x_{ijk} + K a_2 y_0 + \sum_{i=1}^{N} \sum_{k=1}^{K} a_2 x_{ijk} y_i + \sum_{i=0}^{N} \sum_{j=0}^{N} \sum_{k=1}^{K} b x_{ijk} \qquad (11)$$

Subject to

$$\sum_{i=0}^{N} \sum_{k=1}^{K} x_{ijk} = 1 \qquad\qquad for \ i \neq j \ and \ \ j \in \{1,2,...., \ N\} \qquad\qquad (12)$$

$$\sum_{j=0}^{N} \sum_{k=1}^{K} x_{ijk} = 1 \qquad\qquad for \ i \neq j \ and \ \ i \in \{1,2,...., \ N\} \qquad\qquad (13)$$

$$\sum_{j=1}^{N} x_{0jk} = 1 \qquad\qquad for \ \ k \in \{1,2,...., \ K\} \qquad\qquad (14)$$

$$\sum_{i=1}^{N} x_{i0k} = 1 \qquad\qquad for \ \ k \in \{1,2,...., \ K\} \qquad\qquad (15)$$

$$\sum_{i=0}^{N} x_{ijk} = \sum_{i=0}^{N} x_{jik} \qquad\qquad for \ \ k \in \{1,2,...., \ K\} \ and \ \ j \in \{1,2,...., \ N\} \qquad\qquad (16)$$

$$y_0 = Min \ \_ \ load \qquad\qquad\qquad\qquad (17)$$

$$y_i - y_j \geq d_i - M \ (1 - x_{ijk}) \qquad for \ \ i \in \{1,..., \ N\}, \ j \in \{0,...., \ N\} \ and \ \ k \in \{1,..., \ K\} \qquad (18)$$

$$Min \ \_ \ load \ \leq y_i \leq Max \ \_ \ load \qquad for \ \ i \in \{0,..., \ N\} \qquad\qquad (19)$$

$$y_i \geq 0 \qquad\qquad\qquad\qquad (20)$$

$$x_{ijk} \in \{0,1\} \qquad\qquad\qquad\qquad (21)$$

The objective function (11) seeks to minimize fuel consumption related to distance and vehicle load. Constraints (12) and (13) ensure that each customer is visited exactly at once. Constraints (14) and (15) ensure the depart from and return to the depot node. Constraint (16) ensures flow balance for each node. Constraint (17) indicates, vehicles are empty at the depot node. Constraints (18) and (19) restrict the load of vehicles at the visited customer and also eliminate the sub tours. Constraints (20) ensure y_i are nonnegative and constraints (21) define the binary variables x_{ijk}. The objective function (11) is bilinear. In this study, it is observed that, y_i is nonnegative only if $x_{ijk} = 1$. Therefore, the bilinear objective function is converted into the equivalent linear form (22).

$$Min \ z = \sum_{i=0}^{N} \sum_{j=0}^{N} \sum_{k=1}^{K} a_1 c_{ij} x_{ijk} + Ka_2 y_0 + \sum_{i=1}^{N} a_2 y_i + \sum_{i=0}^{N} \sum_{j=0}^{N} \sum_{k=1}^{K} bx_{ijk} \qquad\qquad (22)$$

V. RESULTS AND DISCUSSION

This section presents the results of computatitional applications performed to assess the performance of G-CVRP model. To test the proposed G-CVRP model 15 different instances were generated randomly. The size of the instances generated in this study ranges from 5 to 25 nodes. To better explain the 15 test instances, the charactersitics of the instances, such as; customer number, average load, averagre distance and distance range can be seen in Table 2. The customer number specifies the number of nodes for a given route, the seed number is the code for random number generator, the average load refers the average demand of the customer nodes, the average distance refers to average distance of c_{ij} matrix, and the distance range refers to the difference between max distance and minimum distance of the given route.

Table 2. The characteristics of the 15 instances

Instances	Number of Customers	Average Demand	Average Distance	Distance Range
1	5	50.00	42.66	78.99
2	8	65.63	38.71	64.66
3	10	52.50	51.62	98.61
4	12	62.08	51.05	110.81
5	13	54.62	49.21	89.87
6	15	59.33	43.23	107.24
7	18	57.22	52.02	115.91
8	16	62.81	50.48	99.81
9	15	60.00	51.92	98.77
10	16	74.06	51.11	93.21
11	20	71.75	58.38	125.99
12	22	63.41	53.20	114.82
13	25	54.80	51.58	127.88
14	20	52.75	50.30	94.86
15	21	46.90	52.00	102.32

Proposed G-CVRP model is coded in MPL (Mathematical Programming Language) and solved with GUROBI 5.1.0 solver. Each instance is performed with both CVRP [5] and G-CVRP models and optimum results are obtained. Comparative results of the CVRP and G-CVRP model as to total distance, fuel consumption and vehicle number can be seen in Table 3.There can be differences in vehicle numbers for CVRP and G-CVRP because of the different objective functions of each model. For example the instance set with 10 customer nodes reaches the optimal solution with one vehicle for CVRP and with two vehicles with G-CVRP. It can be concluded from the computational results in Table 3, although the average distance obtained from CVRP (353.81 km) is smaller than average distance obtained from G-CVRP (378.13 km), the average fuel consumption in G-CVRP (25.61 *l*) is less than CVRP (26.79 *l*). In G-CVRP vehicles should serve the customer with large demand earlier in the route. With possible increase in total distance, this approach, reduces the distance weighted average load of vehicles and therefore reduces fuel consumption. This outcome can be clearly seen from the last column of Table 3, such as,

Table 3. Computational Results

Instances	Number of Customers	CVRP			GCVRP			%GAP	
		Total Distance	Fuel Consumption	Number of Vehicles	Total Distance	Fuel Consumption	Number of Vehicles	Distance	Fuel Consumption
1	5	211.93	13.4042	1	211.93	13.4042	1	0.00	0.00
2	8	233.81	15.8502	1	233.81	15.8502	1	0.00	0.00
3	10	299.09	20.8082	1	327.57	20.3957	2	9.52	-1.98
4	12	288.49	21.8091	1	298.69	19.4501	2	3.54	-10.82
5	13	295.34	21.2212	1	295.34	21.2212	1	0.00	0.00
6	15	346.29	29.2608	1	374.88	25.2197	2	8.26	-13.81
7	18	428.51	32.8033	2	438.13	29.8403	2	2.24	-9.03
8	16	369.37	29.3391	2	390.76	26.5671	2	5.79	-9.09
9	15	340.04	26.4800	1	369.77	25.3692	2	8.74	-4.20
10	16	376.43	26.8922	2	376.43	26.8922	2	0.00	0.00
11	20	441.57	33.1009	2	442.41	33.0952	2	0.19	-0.02
12	22	448.64	34.3727	2	460.27	33.6299	2	2.59	-2.16

13	25	474.35	35.9817	2	557.46	34.5884	3	17.52	-3.87
14	20	370.98	29.3911	2	411.24	28.3289	3	10.85	-3.61
15	21	382.37	31.2762	1	483.28	30.3941	3	26.39	-2.82

G-CVRP provides average 4.09 % fuel consumption reduction. The percentage gap refers the percentage difference of CVRP and G-CVRP for each instance. In G-CVRP the model tries to serve the customers with heavier demand earlier and this obtains
the reduction in fuel consumption. It can be concluded from the results that G-CVRP model outperforms CVRP model according to fuel consumption results.

VI.   CONCLUSIONS

In this paper G-CVRP, a variant of widely known VRP, is analyzed and formulated. Differently from classical VRP which tries to minimize total distance of the route, G-CVRP means to minimize total fuel consumption of the route, consequently CO_2 emission of the vehicles in the route. The contribution of this paper is multiple. Initially a computation for calculating fuel consumption considering the vehicle technical specifications, vehicle load and the distance is proposed. Secondly because of the complexity of the fuel consumption calculation, the parameters except load and distance are assumed to be constant and regression analysis is achieved by using Minitab to obtain a mathematical equation between load, distance and fuel consumption. Differently from previous studies in regression analysis both vehicle load and distance are considered in fuel consumption calculation. Finally a mixed integer linear programming model for the G-CVRP is presented and solved optimally with MPL-Gurobi solver. G-CVRP optimization model is validated by various instances. Achieved results show that G-CVRP model provides important reductions in fuel consumption. Results also show that distance minimization does not necessarily mean fuel consumption minimization. Future research can be studying the G-CVRP with larger number of customer instances and also developing heuristic algorithm for quick responses in real-life applications. Proposed fuel consumption equation methodology can be adapted to different VRP formulations, such as pick-up and delivery VRP, capacitated VRP, VRP with time windows and multiple depots VRP.

ACKNOWLEDGEMENT

REFERENCES

[1]   O Jabali, T Van Woensel, AG Kok. Analysis of travel times and CO_2 emissions in time-dependent vehicle routing. Production and Operations Management, 21(6): 1060-1074, 2012.
[2]   G Laporte. The Vehicle routing problem: An overview of exact and approximate algorithms. European Journal of Operations Research, 59: 345-358, 1992.
[3]   Y Xiao, Q Zhao, I Kaku, Y Xu. Development of a fuel consumption optimization model for the capacitated vehicle routing problem. Computers & Operations Research, 39: 1419-1431, 2012.
[4]   L D Bodin. Twenty years of routing and scheduling. Operations Research, 38(4): 571-579, 1990.
[5]   P Toth, D Vigo. The vehicle routing problem. SIAM. Philadelphia: PA; 2002.
[6]   A Hoff, H Andersson, M Christiansen, G Hasle, A Lokketangen. Industrial aspects and literature survey: Fleet composition and routing. Computers & Operations Research, 37: 2041-2061, 2010.
[7]   Y Huang, C Shi, L Zhao, T Van Woensel. A study on carbon reduction in the vehicle routing problem with simultaneous pickups and deliveries. IEEE International Conference on Service Operations and Logistics and Informatics. 8-10 July 2012. Suzhou, China.
[8]   Ö Apaydın, MT Gönüllü. Emission control with route optimization in solid waste collection process: A case study. Sadhana, 33(2): 71-82, 2008.
[9]   Y Kuo. Using simulated annealing to minimize fuel consumption fort the time-dependent vehicle routing problem. Computers & Industrial Engineering, 59: 157-165, 2010.
[10]   T Bektaş, G Laporte. The pollution-routing problem. Transportation Research Part B, 45: 1232-1250, 2011.
[11]   Y Suzuki. A new truck-routing approach for reducing fuel consumption and pollutants emission. Transportation Research Part D, 16: 73-77, 2011.
[12]   E Wygonik, A Goodchild. Evaluating CO_2 emissions, cost, and service quality trade-offs in an urban delivery system case study. IATTS Research, 35: 7-15, 2011.
[13]   S Erdoğan, E Miller-Hooks. A green vehicle routing problem. Transportation Research Part E, 48: 100-114, 2012.
[14]   S Ubeda, FJ Arcelus, J Faulin. Green logistics at Eroski: A case study. International Journal of Production Economics, 131: 44-51, 2011.
[15]   E Demir, T Bektaş, G Laporte. A comparative analysis of several vehicle emission models for road freight transport. Transportation Research Part D, 16: 347-357, 2011.
[16]   E Demir, T Bektaş, G Laporte. An adaptive large neighborhood search heuristic for the Pollution-Routing Problem. European Journal of Operations Research,   223: 346-359, 2012.
[17]   Y Liu, GE Helfand. The alternative motor fuels act, alternative-fuel vehicles, and greenhouse gas emissions. Transportation Research Part A, 43: 755-764, 2009.

[18]   HH Braess, U Seiffert. Handbook of Automotive Engineering. SAE International. Pennsylvania USA; 2005.
[19]   P Toth, D Vigo. Exact algorithms for vehicle routing. In: Crainic T, Laporte G, editors. Fleet management and logistics, Boston: Kluwer Academic Publishers; p. 1–31, 1998.

Lattice Points On The Homogeneous Cubic Equation With Four Unknowns

$$x^2 - xy + y^2 + 3w^2 = 7z^3$$

M.A.Gopalan[1], V.Sangeetha[2] , Manju Somanath[3]

[1] *Professor, Department of Mathematics, Srimathi Indhira Gandhi College,Trichy-*
[2] *Assistant Professor, Department of Mathematics, National College,Trichy-1.*
3Assistant Professor, Department of Mathematics, National College,Trichy-1.

ABSTRACT
The homogeneous cubic equation with four unknowns represented by $x^2 - xy + y^2 + 3w^2 = 7z^3$ is analyzed for its patterns of non zero distinct integral solutions.Three different patterns of solutions are presented. A few interesting relations between the solutions and special numbers ,namely, polygonal numbers,pyramidal numbers,centered pyramidal numbers,star numbers, nasty numbers,dodecahedral number,rhombic dodecahedral number and prism numbers are exhibited.

KEY WORDS: Homogeneous cubic,Lattice Points,Integral solutions. MSC2000 Mathmatics Subject Classification Number 11D25.

Notations

$t_{m,n}$= Polygonal number of rank n with sides m.

P_n^m = Three dimensional figurate number of rank n with sides m

cP_n^m = Three dimensional centered figurate number of rank n with sides m

S_n= Star number

$D(n)$ = Dodecahedral number of rank n

$RD(n)$ = Rhombic dodecahedral number of rank n

PCS_m^n = Prism number of rank n with sides m.

I.   INTRODUCTION

The Cubic Equation offers an unlimited field for research because of their variety [1,2]. In particular, one may refer [3-13] for cubic equation with four unknowns. This   communication   concerns with   yet another   interesting   equation $x^2 - xy + y^2 + 3w^2 = 7z^3$ representing a homogeneous cubic with four unknowns for determining its infinitely many non-zero integral solutions. Also a few interesting relations among the solutions have been presented.

II.   Method of Analysis

The homogeneous cubic represented by the cubic equation is

$$x^2 - xy + y^2 + 3w^2 = 7z^3 \qquad\qquad (1)$$

It is observed that (1) is satisfied by infinitely many non-zero distinct integral solutions. For the sake of clear understanding, we present below different patterns of solutions to (1).

2.1 Pattern - 1

Introducing the linear transformations

$$x = u + v \; ; \; y = u - v \; ; \; w = v \qquad\qquad (2)$$

in (1), it is written as

$$u^2 + 6v^2 = 7z^3 \qquad\qquad (3)$$

Assume $z = a^2 + 6b^2$ \qquad\qquad (4)

Write 7 as 7= $(1 + i\sqrt{6})(1 - i\sqrt{6})$ \qquad\qquad (5)

Using (4) and (5) in (3) and applying the method of factorization, define

$$u + i\sqrt{6}v = (a + i\sqrt{6}b)^3 (1 + i\sqrt{6})$$

Equating the real and imaginary parts, we get

$$u = a^3 - 18ab^2 - 18a^2b + 36b^3$$
$$v = a^3 - 18ab^2 + 3a^2b - 6b^3$$
$$\left.\right\} \qquad (6)$$

Using (6) in (2), the integral solutions to (1) is obtained as

$$x(a,b) = 2a^3 - 36ab^2 - 15a^2b + 30b^3$$
$$y(a,b) = -21a^2b + 42b^3$$
$$z(a,b) = a^2 + 6b^2$$
$$w(a,b) = a^3 - 18ab^2 + 3a^2b - 6b^3$$

A few properties of the above solutions are

1. $2x(a(a+1),1) - 6P_{a(a+1)}^6 + 66t_{3,a(a+1)} + 76t_{3,a} = 60$.
2. $.x(1,b) - 6P_b^{30} - 18t_{3,b} + 45t_{4,b} = 2$.
3. $y(1,b) - x(1,b) - 6P_b^{14} - 6t_{3,b} - 30t_{4,b} = -2$.
4. $y(1,b) + z(1,b) - 6P_b^{44} - 6t_{3,b} \equiv 1 \ (mod \ 15)$.
5. $x(1,b) + w(1,b) - 6P_b^{24} - 12t_{3,b} + 60t_{4,b} = 3$.

In addition to (5), 7 can also be written as

$$7 = \frac{(13 + i\sqrt{6})(13 - i\sqrt{6})}{25} \qquad and \qquad 7 = \frac{(29 + i\sqrt{6})(29 - i\sqrt{6})}{121} \qquad (7)$$

Applying the procedure similar to pattern 1, we obtain the corresponding integer solutions to (1) which are given below respectively .

$$x(A,B) = 350A^3 - 6300AB^2 + 525A^2B - 1050B^3$$
$$y(A,B) = 300A^3 - 5400AB^2 - 1425A^2B + 2850B^3$$
$$z(A,B) = 25A^2 + 150B^2$$
$$w(A,B) = 25A^3 - 450AB^2 + 975A^2B - 1950B^3$$

and

$$x(A,B) = 3630A^3 - 65340AB^2 + 8349A^2B - 16698B^3$$
$$y(A,B) = 3388A^3 - 60984AB^2 - 12705A^2B + 25410B^3$$
$$z(A,B) = 121A^2 + 726B^2$$
$$w(A,B) = 121A^3 - 2178AB^2 + 10527A^2B - 21054B^3$$

2.2 Pattern - 2

Introducing the linear transformations

$$x = u + v \ ; \ y = u - v \ ; \ w = u \qquad (8)$$

in (1), it is written as

$$4u^2 + 3v^2 = 7z^3 \qquad (9)$$

Assume $z = 4a^2 + 3b^2$ $\qquad (10)$

Write 7 as $\quad 7 = \frac{(5 + i\sqrt{3})(5 - i\sqrt{3})}{4}$ $\qquad (11)$

Using (10) and (11) in (9) and applying the method of factorization, define

$$2u + i\sqrt{3}v = \left(2a + i\sqrt{3}b\right)^3 \frac{5 + i\sqrt{3}}{2}$$

Equating the real and imaginary parts, we get

$$u = 10a^3 - \frac{45}{2}ab^2 - 9a^2b + \frac{9}{4} \quad and \quad v = 4a^3 - 9ab^2 + 30a^2b - \frac{15}{2}b^3$$

As our interest centers on finding integer solution we choose suitable values for *u* and *v*.
Letting $a = A$, $b = 2B$, we get

$$u = 10A^3 - 90AB^2 - 18A^2B + 18B^3$$
$$v = 4A^3 - 36AB^2 + 60A^2B - 60B^3$$
$$\left.\right\} \qquad (12)$$

Using (12) in (8), the integral solutions to (1) is obtained as

$$x(A,B) = 14A^3 - 126AB^2 + 42A^2B - 42B^3$$
$$y(A,B) = 6A^3 - 54AB^2 - 78A^2B + 78B^3$$
$$z(A,B) = 4A^2 + 12B^2$$
$$w(A,B) = 10A^3 - 90AB^2 - 18A^2B + 18B^3$$

Some interesting properties of the above solutions are

1. $2z(1,B) - 8$ is a Nasty Number.
2. $w(1,B) - 6P_B^{20} + 6t_{3,b} + 90t_{4,b} = 10$.

3. $z(A,1) + w(A,1) - 6P_A^{12} + 34t_{3,A} - t_{4,3A-1} + 9t_{4,A} = -91.$

4. $w(A(A+1),1) - 2D(A(A+1)) + 6P_{A(A+1)}^5 + 192t_{3,A} - 19$ is a Rhombic dodecahedral number.

5. $x(A,1) - y(A,1) - 3CP_A^{16} - 240t_{3,A} \equiv 67 \ (mod \ 187).$

6. $y(1,B) - 2PCS_B^{18} + 144t_{3,B} \equiv 0 \ (mod \ 2).$

In addition,by writing 7 as $7=(2 + i\sqrt{3})(2 - i\sqrt{3})$,we obtain the integer solutions to (1) as below

$$x(A,B) = 16A^3 - 144AB^2 + 12A^2B - 12B^3$$
$$y(A,B) = 84B^3 - 84A^2B$$
$$z(A,B) = 4A^2 + 12B^2$$
$$w(A,B) = 8A^3 - 72AB^2 - 36A^2B + 36B^3$$

III. CONCLUSION

To conclude, one may search for other patterns of solutions and their corresponding properties.

REFERENCES

[1]   L.E.Dickson, *History of Theory of Numbers*, Vol.2, Chelsea Publishing Company,New York, 1952.

[2]   L.J.Mordell, *Diophantine Equations*, Academic Press, London, 1969.

[3]   M.A.Gopalan, S.Premalatha, *Integral Solutions of* $(x+y)(xy+w^2) = 2(k^2+1)z^2$, Bulletin of Pure and Applied Sciences, (2009) Vol.29E, No:2, 197-202.

[4]   M.A.Gopalan, V.Pandichelvi, *Remarkable Solutions on the cubic equation with four unknowns* $x^3 + y^3 + z^3 = 28(x+y+z)w^2$, Antarctica J.of Math,(2010)   Vol.7,No.4 ,393-401.

[5]   M.A.Gopalan,B.Sivagami, *Integral Solutions of homogeneous cubic equation with four unknowns* $x^3 + y^3 + z^3 = 3xyz + 2(x+y)w^3$,Impact J.Sci.Tech., (2010)   Vol.4,No.3, 53-60.

[6]   M.A.Gopalan,S.Premalatha *On the cubic Diophantine Equation with four unknowns* $(x-y)(xy-w^2) = 2(n^2+2n)z^3$, International Journal of Mathematical Sciences, (2010),Vol.9, No.1-2, 171-175.

[7]   M.A.Gopalan,J.Kaliga Rani, *Integral solutions of* $x^3 + y^3 + (x+y)xy = z^3 + w^3 + (z+w)zw$,Bulletin of Pure and Applied Sciences, (2010),Vol.29E,No:1,   169-173.

[8]   M.A.Gopalan, S.Premalatha, *Integral Solutions of* $(x+y)(xy+w^2) = 2(k+1)z^2$, The Global Journal of Applied Mathematics and Mathematical Sciences, (2010),Vol.3, No.1-2,51-55.

[9]   M.A.Gopalan,S.Vidhyalakshmi and S.Mallika,*Observation on cubic equation with four unknowns* $xy + 2z^2 = w^3$,The Global Journal of Mathematics and Mathematical Sciences,(2012) Vol.2,No:1-2, 69-74.

[10]   M.A.Gopalan,S.Vidhyalakshmi and S.Mallika,*Observation on cubic equation with four unknowns* $2(x^3 + y^3) = z^3 + w^3(x+y)$, IJAMP,(2012) Vol.4,No:2 July-December ,103-107.

[11]   M.A.Gopalan,S.Vidhyalakshmi and G.Sumathi,*On the homogeneous cubic equation with four unknowns* $x^3 + y^3 = 14z^3 - 3w^2(x+y)$,Discovery(2012) Vol.2,No.4, October ,17-19

[12]   M.A.Gopalan and K.Geetha, ,*Observation on cubic equation with four unknowns* $x^3 + y^3 + xy(x+y) = z^3 + 2(x+y)w^3$, International Journal of Pure and Applied   Mathematical Sciences, (2013),Vol.6, No.1, 25-30.

[13]   M.A.Gopalan,V.Sangeetha and Manju Somanath,*Lattice Points on the Homogeneous cubic equation with four unknowns* $(x+y)(xy+w^2) = (k^2-1)z^3, k > 1$,   International Journal of Pure and Applied Mathematical Sciences,(2013),Vol.6,No.2,241-246.

Research on Cloud Computing

[1,] Done VinodKumar, [2,] Thudum Venkatesh, [3,] Kondabathula Durga charan,
[4,] Chandrasekhar T

[1,2,3,4,] M. Tech in Computational Engineering, RGUKT

ABSTRACT

This paper is an introduction to the terms, characteristics, and services associated with network-based computing, commonly referred to as cloud computing. Cloud Computing is a technology that uses the network and central remote servers to maintain data and applications. Cloud computing allows consumers and businesses to use applications without installation and access their personal files at any computer with network access. This technology allows for much more efficient computing by centralizing data storage, processing and bandwidth. The primary business service models being deployed (such as software, platform, and infrastructure as a service) and common deployment models employed by service providers and users to use and maintain the cloud services (such as the private, public, community, and hybrid clouds) are discussed.

KEYWORDS: *Architecture, Shared infrastructure, Network access, managed metering, Service models, Deployment models, Software as a Service (SAAS), Platform as a Service (PAAS), and Infrastructure as a Service (IAAS).*

I. INTRODUCTION

Cloud computing describes the approach, abstracted IT infrastructures (computing capacity, data storage, network capacity, or even finished software) dynamically adapted to the needs of a network to provide. From the user point of view, provided abstracted IT infrastructure seems distant and obscure, hidden in a "cloud" to happen. Supply and utilization of these services take place exclusively via defined technical interfaces and protocols. The range of cloud computing as part of services offered covers the complete spectrum of information technology and includes, among other things, infrastructure (such as processing power, disk space), platforms and software. In simplified terms one can follow the concept can be described as: A part of the IT environment (in this context, such as hardware such as data center, data storage and software) is on the user side not operated or be provided locally, but with one or more providers hired as a service usually located geographically far away. The applications and data are no longer on the local machine or in the corporate data center, but in the (metaphorical) cloud. The design element of an abstract outline of clouds in network diagrams often used to represent an unspecified part of the Internet.

The access to the remote systems of a network, such as that of the Internet is in use. But there are also companies in the context of so-called private clouds, where provision of a corporate intranet is done. Most providers of cloud solutions leverage the pooling effects that arise from the sharing of resources for their business model.

II. ARCHITECTURE

Cloud reference architecture model, which portrays the architectural layers, connected via APIs and repositories. If we study the framework more closely, the following aspects can be articulated. At the foundation of this framework is the data center technology architecture, which consists of three salient blocks of network, compute, and storage. This layer hosts all the services that are delivered to a cloud consumer or subscriber. This layer will be discussed in more detail in later sections of this paper. The next important layer is the security layer. The key takeaway in this layer is that security is blanketed as an end-to-end architecture across all aspects of the framework. Security is considered as one of the key challenges to be solved in a cloud framework; hence, it has to be accounted for in a comprehensive sense. This layer will be discussed in more detail in later sections of this paper.
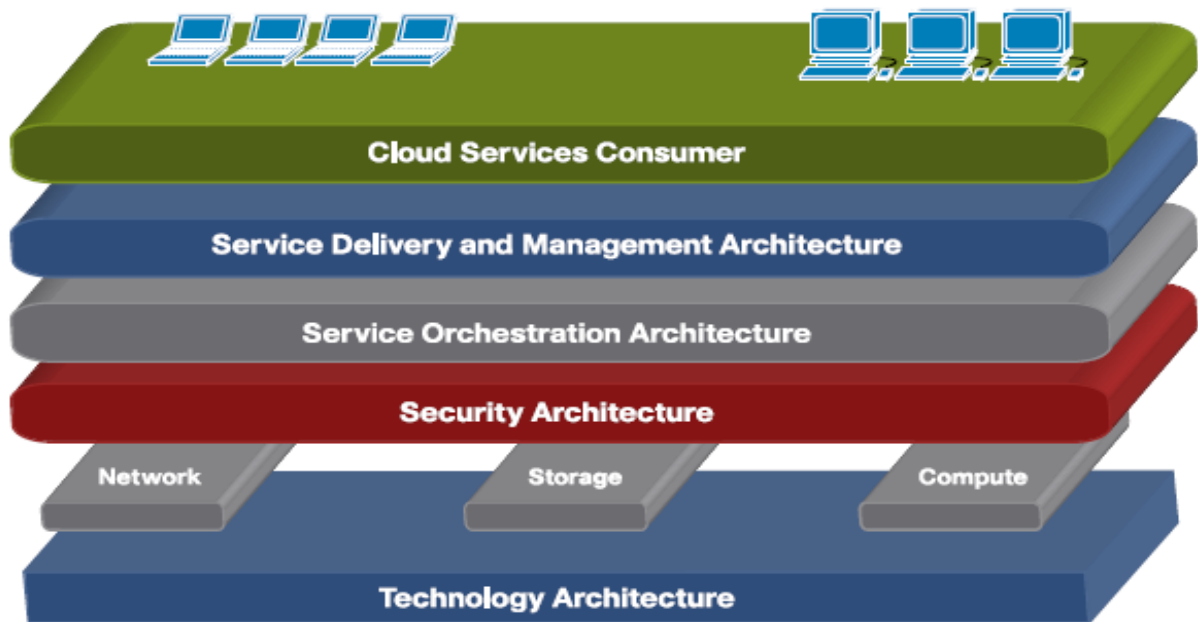
Fig1.   Cloud Reference Architecture Framework

Following the technology and security layer is the Service Orchestration layer, which is implemented with configuration repository enablers. The configuration repository stores key information such as service catalogue, asset inventory, and resource-to-service mappings. This layer is an important layer because it maps the technology components to the service components and serves as a reference point during service provisioning. The service orchestration layer is the "glue" that integrates the lower layers to create a service for delivery.The next layer is also where infrastructure and service management function take place.The topmost layer is the consumer-facing layer, usually exposed via a portal-like solution.This is the layer where service is defined, requested, and managed by the consumer.

Let's walk through a use case scenario where this framework is utilized.

1   Consumer logs on to a cloud portal and verifies/updates credentials and information.
2   Based on the consumer entitlement, a selected a set of services are identified and presented for definition.
3   The end user selects the service for consumptions and triggers a service-provisioning request.
4   Resources are marked as reserved for service, and a new request is created for services provisioning.
5   The individual domains of compute, network, and storage are configured and provisioned, with requested security and service-level agreements (SLAs), for service delivery.

III.   CHARACTERISTICS
Cloud computing has a variety of characteristics, with the main ones being.

3.1 Shared Infrastructure
Shared infrastructure uses a virtualized software model, enabling the sharing of physical services, storage, and networking capabilities. The cloud infrastructure, regardless of deployment model, seeks to make the most of the available infrastructure across a number of users.

3.2 Dynamic Provisioning
Dynamic provisioning allows for the provision of services based on current demand requirements. This is done automatically Using software automation, enabling the expansion and contraction of service capability, as needed. This dynamic scaling needs to be done while maintaining high levels of reliability and security.

3.3 Network Access

Network access needs to be accessed across the internet from a broad range of devices such as PCs, laptops, and mobile devices, using standards-based APIs (for example, ones based on HTTP). Deployments of services in the cloud include everything from using business applications to the latest application on the newest smart phones.

3.4 Managed Metering

Managed metering uses metering for managing and optimizing the service and to provide reporting and billing information. In this way, consumers are billed for services according to how much they have actually used during the billing period. In short, cloud computing allows for the sharing and scalable deployment of services, as needed, from almost any location, and for which the customer can be billed based on actual usage.

IV.   SERVICE MODELS

Cloud computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the data centers that provide those services. The services themselves have long been referred to as Software as a Service (SaaS). Some vendors use terms such as IaaS (Infrastructure as a Service) and PaaS (Platform as a Service) to describe their products, but we eschew these because accepted definitions for them still vary widely.

Service Class	Main Access & Management Tool	Service content
SaaS	Web Browser	**Cloud Applications** Social networks, Office suites, CRM, Video processing
PaaS	Cloud Development Environment	**Cloud Platform** Programming languages, Frameworks, Mashups editors, Structured data
IaaS	Virtual Infrastructure Manager	**Cloud Infrastructure** Compute Servers, Data Storage, Firewall, Load Balancer

Fig2.   The cloud computing stack

Once a cloud is established, how its cloud computing services are deployed in terms of business models can differ depending on requirements. The primary service models being deployed (see Figure 2) are commonly known as.

4.1 Software as a Service (SaaS)

Consumers purchase the ability to access and use an application or service that is hosted in the cloud. A benchmark example of this is Salesforce.com, as discussed previously, where necessary information for the interaction between the consumer and the service is hosted as part of the service in the cloud. Also, Microsoft is expanding its involvement in this area, and as part of the cloud computing option for Microsoft Office 2010, its Office Web Apps are available to Office volume licensing customers and Office Web App subscriptions through its cloud-based on-line services.

4.2 Platform as a Service (PaaS)

Consumers purchase access to the platforms, enabling them to deploy their own software and applications in the cloud. The operating systems and network access are not managed by the consumer, and there might be constraints as to which applications can be deployed.

4.3 Infrastructure as a Service (IaaS)

Consumers control and manage the systems in terms of the operating systems, applications, storage, and network connectivity, but they do not control the cloud infrastructure. Also known are the various subsets of these models that may be related to a particular industry or market. Communications as a Service (CaaS) is one

such subset model used to describe hosted IP telephony services. Along with the move to CaaS is a shift to more IP-centric communications and more SIP trucking deployments. With IP and SIP in place, it can be as easy to have the PBX in the cloud as it is to have it on the premise. In this context, CaaS could be seen as a subset of SaaS.

V.   DEPLOYMENT MODELS

Deploying cloud computing can differ depending on requirements, and the following four deployment models have been identified, each with specific characteristics that support the needs of the services and users of the clouds in particular ways. Although cloud computing has emerged mainly from the appearance of public computing utilities, other deployment models, with variations in physical location and distribution, have been adopted. In this sense, regardless of its service class, a cloud can be classified as public, private, community, or hybrid based on model of deployment as shown in Figure
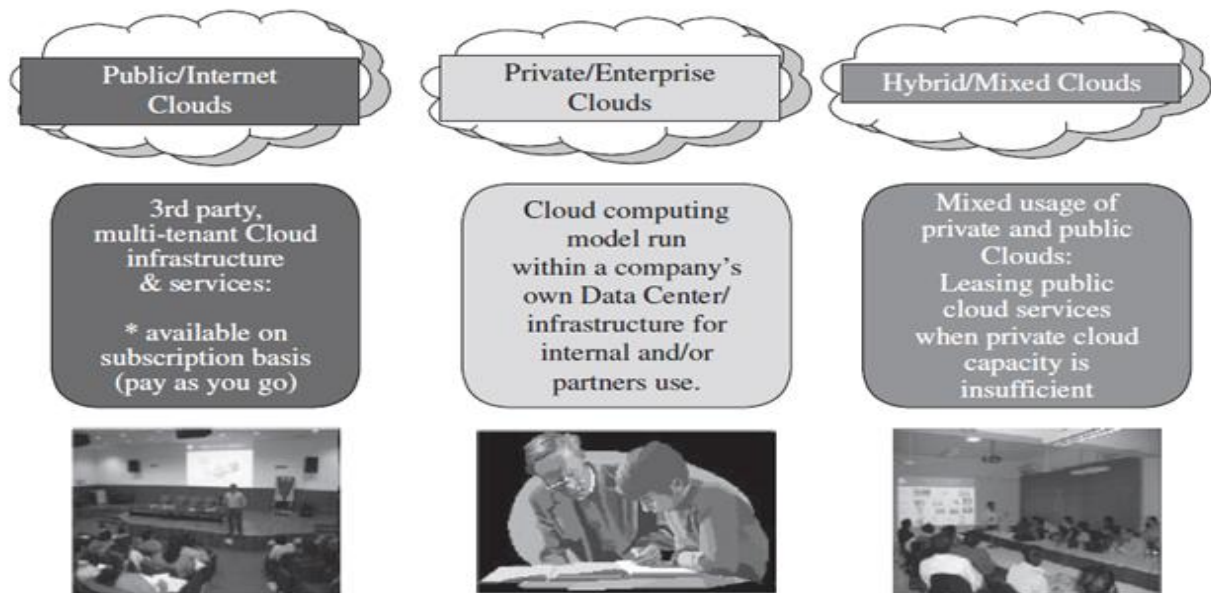
Fig3.   Types of clouds based on deployment models.

5.1 Private Cloud

The cloud infrastructure has been deployed, and is maintained and operated for a specific organization. The operation may be in-house or with a third party on the premises.

5.2 Public Cloud

The cloud infrastructure is available to the public on a commercial basis by a cloud service provider. This enables a consumer to develop and deploy a service in the cloud with very little financial outlay compared to the capital expenditure requirements normally associated with other deployment options.

5.3 Hybrid Cloud

The cloud infrastructure consists of a number of clouds of any type, but the clouds have the ability through their interfaces to allow data and/or applications to be moved from one cloud to another. This can be a combination of private and public clouds that support the requirement to retain some data in an organization, and also the need to offer services in the cloud.

VI.   UNDERLYING TECHNOLOGIES

Virtualization, Clustering and Grid computing are simply some types of underlying technologies for implementing cloud computing.

6.1 Virtualization

Internet, it is a large collection of networks where resources are globally networked, In internet cloud computing plays a major role In order to share the data and one of the important technology in the cloud computing is virtualization. Mainly it is used to maintain the collection IT resources which are used by the cloud

providers. The main aim of the virtualization is ability to run the multiple operating systems on a single machine buy sharing all the resources that belong to the hard ware. In this paper our main goal is to provide the basic knowledge about the virtualization technology in cloud computing and how it acts in the cloud computing environment. And we also discuss about, how to maintain the virtualized environment with the optimized resources.

Fig4. Server Virtualization

In virtualization, cloud back end layer which holds the key access to physical resources are bind with hyper-visor layer and the technique of virtualization is performed at that time, virtualization breaks the one physical resource into the multiple virtual resources and controls the access limit and utilization to single physical resource.These multiple virtual resources are then allowed to be shared across different consumers. This brings the optimum resource utilization.

6.2 Clustering
        On the other hand if demand increases by the consumer where he requires more the potential of existing physical resource and does not really have the time to do the remodeling of resource sharing, cloud comes with benefits of scalability, now to use this effectively engineers uses the concept of sharing called clustering, this is grouping or binding the two or more physical resources in such a way that pool of availability never dries out. On top it makes it look like a single virtual entity to design to satisfy the operational requirement.
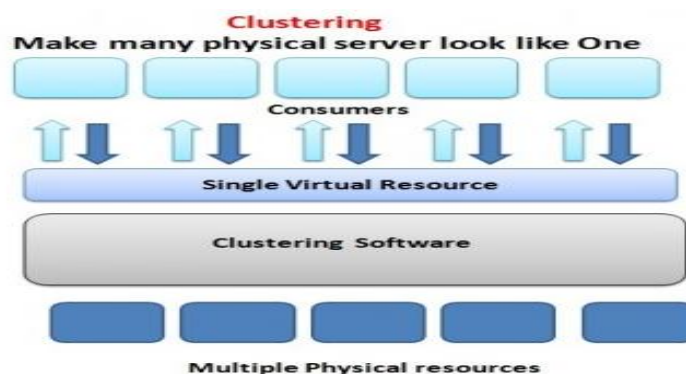
Fig5. Clustering

6.3 Grid Computing

Grid computing is the federation of computer resources from multiple locations to reach a common goal. The grid can be thought of as a distributed system with non-interactive workloads that involve a large number of files. What distinguishes grid computing from conventional high performance computing systems such as cluster computing is that grids tend to be more loosely coupled, heterogeneous, and geographically dispersed. Although a single grid can be dedicated to a particular application, commonly a grid is used for a variety of purposes. Grids are often constructed with general-purpose grid middleware software libraries. Internet, it is a large collection of networks where resources are globally networked, In internet cloud computing plays a major role In order to share the data and one of the important technology in the cloud computing is virtualization. Mainly it is used to maintain the collection IT resources which are used by the cloud providers. The mainaim of the virtualization is ability to run the multiple operating systems on a single machine buy sharing all the resources that belong to the hard ware. In this paper our main goal is to provide the basic knowledge about the virtualization technology in cloud computing and how it acts in the cloud computing environment. And we also discuss about, how to maintain the virtualized environment with the optimized resources.

VII.   CONCLUSION

Among the most popular cloud services now are social networking sites (the 500 million people using Face book are being social in the cloud), webmail services like Hotmail and Yahoo mail, micro blogging and blogging services such as Twitter and Word Press, video-sharing sites like YouTube, picture-sharing sites such as Flick, document and applications sites like Google Docs, social-bookmarking sites like Delicious, business sites like eBay, and ranking, rating and commenting sites such as Yelp and Trip Advisor.A simple example of cloud computing is Yahoo email, Gmail, or Hotmail etc. All you need is just an internet connection and you can start sending emails. The server and email management software is all on the cloud (internet) and is totally managed by the cloud service provider Yahoo, Google etc. The consumer gets to use the software alone and enjoy the benefits.
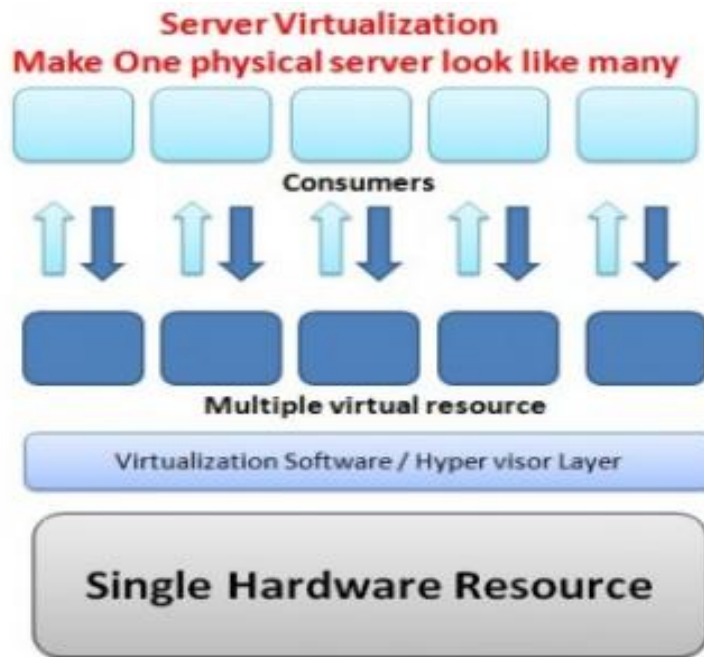
REFERENCES

[1].   Wikipedia, "Cloud Computing" http://en.wikipedia.org/wiki/Cloud_computing, Oct 12, 2012
          http://simple.wikipedia.org/wiki/Cloud_computing
[2].   R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, Cloud computingand emerging IT platforms: Vision, hype, and reality for delivering computing asthe 5th utility, Future Generation Computer Systems, 25:599_616, 2009.
[3].   L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, A break in theclouds: Towards a cloud definition, SIGCOMM Computer CommunicationsReview, 39:50_55, 2009.
[4].   National Institute of Standards and Technology, "The NIST Definition of Cloud Computing"
[5].   How Stuff Works, "Cloud Computing Working",
[6].   McKinsey & Co., Clearing the Air on Cloud Computing, Technical Report, 2009.On Cloud Computing, "Cloud computing Technology/cloud computing architecture", http://www.oncloudcomputing.com/en/2009/09/cloud-computing-architecture-cloud-computing-technology/, Feb 26, 2012.
[7].   "Cloud computing-The New Frontier of Internet Computing", Science Direct, the MITRE Corporation, George Pallis.
[8].   How does the cloud computing works and technologies behind it
          http://www.diceitwise.com/how-does-cloud-computing-work-and-technology-behind-it/#.UcCpfPL_zxq
[9].   Advance in Cloud. Research in Future of Cloud Computing
[10].   The Future of Cloud Computing by Janna Quitney Anderson, Elon University and   Lee Rainie, Pew Internet & American Life Project June 11, 2010
[11].   International Business Machines Corp., An architectural blueprint for autonomiccomputing, White Paper Fourth Edition, 2006.
[12].   R. P. Goldberg, Survey of virtual machine research, IEEE Computer, 7(6):34_45, 1974.
[13].   IBM, "Cloud Computing", http://www.ibm.com/cloud-computing/us/en/, March 30, 2012.
[14].   "Cloud computing for Beginners", http://www.techno-pulse.com/
[15].   Cloud computing Web based Application that change the way you work and collaborative online
[16].   A brief guide to cloud computing by Christopher Barnett
[17].   Blau, D. Neumann, C. Weinhardt, and S. Lamparter, Planning and pricing ofservice mashups, in Proceedings of the 2008 10th IEEE Conference on E-commerce Technology and the Fifth IEEE Conference on Enterprise Computing, E-Commerce and E-Services, Crystal City, Washington, DC, 2008, pp.19_26.
[18].   Sotomayor, R. S. Montero, I. M. Llorente, and I. Foster, Virtual infrastructure Management in private and hybrid clouds, IEEE Internet Computing, 13(5):14_22, September/October, 2009.

A Review Of Laminar Burning Velocity Of Gases And Liquid Fuels.

Vaishali Katre[1], S. K. Bhele[2]

[1, 2]Department of Mechanical Engineering, Kavikulguru Institute of Technology and Science Ramtek, India

ABSTRACT:

The laminar burning velocity is a fundamental property of a fuel that affects many aspects of its combustion behaviour. Experimental values are required to validate kinetic simulations, and also to provide input for models of flashback, minimum ignition energy and turbulent combustion. The laminar burning velocity is one of the most essential parameters for analysis and performance predictions of various combustion engines. It is the velocity, relative to the unburned gas, with which a plane, one-dimensional flame front travels along the normal to its surface. The majority of turbulent combustion models require knowledge of laminar burning velocity of the fuel-air mixture as a function of the mixture strength. Also reliable experimental data are needed in order to test and calibrate thermo-kinetic combustion models which have been quite successful for combustion predictions of simple hydrocarbon fuels. There are different methods to determine the burning velocity such as Heat flux burner method, Flat flame burner method, Bunsen burner method, Slot burner technique, Counter flow diffusion flow, Soap bubble technique, and Tube propagating method.

KEYWORDS: *Burning velocity, Bunsen burner, Flame front, Flame speed.*

I.   INTRODUCTION

The laminar burning velocity is one of the fundamental properties of a reacting premixed mixture and its reliable data are constantly needed for combustion applications. So far, several techniques for measuring the one-dimensional laminar burning velocity have been used, and for a wide range of temperature, pressure, and fuel rather accurate measurements have been obtained by employing flat or curved flames in stagnation flow, propagating spherical flames in combustion vessel or flat flames stabilized on burner. With all those measurement technique proper care could be taken to remove the effect of flame stretch either during experimentation or through further data processing.

1.1 Importance of laminar burning velocity

Laminar burning velocity is an important parameter of a combustible mixture as it contains fundamental information regarding reactivity, diffusivity, and exothermicity. Its accurate knowledge is essential for engine design, modeling of turbulent combustion, and validation of chemical kinetic mechanisms. In addition, the determination of burning velocity is very important for the calculations used in explosion protection and fuel tank venting. Burning velocity is defined as the linear velocity of the flame front normal to itself relative to unburned gas, or as the volume of unburned gas consumed per unit time divided by the area of the flame front in which that volume is consumed. Laminar burning velocity is highly useful for modeling turbulent burning velocity. Turbulent flow occurs when a fluid undergoes irregular fluctuations and mixing. Laminar flow is defined as the flow which travels smoothly in regular paths or layers.

1.2 Flame front propagation

For efficient combustion the rate of propagation of the flame front within the cylinder is quite critical. The two important factors which determine the rate of movement of the flame front across the combustion chamber are the reaction rate and the transposition rate. The reaction rate is the result of a purely chemical combination process in which the flame eats its way into the unburned charge. The transposition rate is due to the physical movement of the flame front relative to the cylinder wall and is also the result of the pressure differential between the burning gases and the unburnt gases in the combustion chamber.
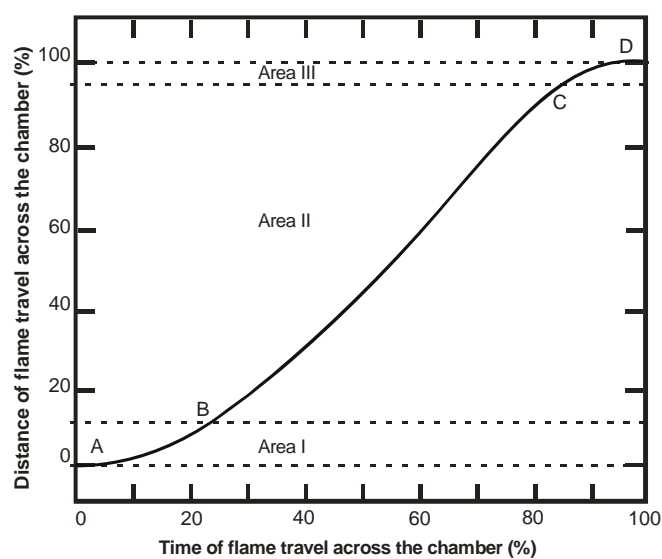
Figure 1: Details of flame travel

Figure shows the rate of flame propagation. In area 1, (A→B), the flame front progresses relatively slowly due to a low transposition rate and low turbulence. The transposition of the flame front is very little since there is a comparatively small mass of charge burned at the start. The low reaction rate plays a dominant role resulting in a slow advance of the flame. Also, since the spark plug is to be necessarily located in a quiescent layer of gas that is close to the cylinder wall, the lack of turbulence reduces the reaction rate and hence the fame speed. As the flame front leaves the quiescent zone and proceeds into more turbulence areas (area2) where it consumes a greater mass of mixture, it progresses more rapidly and at a constant rate (B→C) as shown in figure. The volume of unburned charge is very much less towards the end of flame travel and so the transposition rate again becomes negligible thereby reducing the flame speed. The reaction rate is also reduced again since the flame is entering a zone area (area 3) of relatively low turbulence (C→D) in figure.

1.3 Factor influencing the flame speed

The study of factors which affect the velocity of flame propagation is important since the flame velocity influences the rate of pressure rise in the cylinder and it is related to certain types of abnormal combustion that occur in spark-ignition engines. There are several factors which affect the flame speed, to a varying degree, the most important being the turbulence and the fuel-air ratio. Details of various factors that affect the flame speed are discussed below.

1.3.1 Turbulence

The flame speed is quite low in non-turbulent mixtures and increases with increasing turbulence. This is mainly due to the additional physical intermingling of the burning and unburned particles at the flame front which expedites reaction by increasing the rate of contact. The turbulence in the incoming mixture is generated during the admission of fuel- air mixture through comparatively narrow sections of the intake pipe, valve openings etc., in the suction stroke. Turbulence which is supposed to consist of many minute swirls appears to increase the rate of reaction and produce a higher flame speed than that made up of larger and fewer swirls. A suitable design of the combustion chamber which involves the geometry of cylinder head and piston crown increases the turbulence during the compression stroke.

1.3.2 Fuel-air ratio

The fuel air ratio has a very significant influence on the flame speed. The highest flame velocities are obtained with somewhat richer mixture. When the mixture is made leaner or richer the flame speed decreases. Less thermal energy is released in the case of lean mixtures resulting in lower flame temperature. Very rich mixtures lead to incomplete combustion which results again in the release of less thermal energy. 1.3.3 Temperature and pressure Flame speed increases with an increase in intake temperature and pressure. Higher initial pressure and temperature may help to form a better homogeneous air-vapour mixture which helps in increasing the flame speed. This is possible because of an overall increase in the density of the charge.

1.3.4 Compression ratio

A higher compression ratio increases the pressure and temperature of the working mixture which reduce the initial preparation phase of combustion and hence less ignition advance is needed. High pressures and temperatures of the compressed mixture also speed up the second phase of combustion. Increase compression ratio reduces the clearance volume and therefore increases the density of the cylinder gases during burning. This increases the peak pressure and temperature and total combustion duration is reduced. Thus engines having higher compression ratios have higher flame speeds. According to S.Y. Liao et al carried out the study on Determination of the laminar burning velocities for mixtures of ethanol and air at elevated temperatures. It has measured the laminar burning velocities for ethanol-air premixed flames at various temperature and equivalence ratio. The flames are analyzed to estimate flame size, consequently, the flame speeds are derived from the variations of the flame size against the time elapsed. It was studied the effects of the fuel/air equivalence ratio, initial temperature and pressure on the laminar flame propagation. It was conducted the premixed laminar combustion of ethanol-air mixture experimentally in a closed combustion bomb. And it was found the laminar burning velocity 58.3 cm/s at normal pressure of 0.1 MPa and temperature of 358 K.Xuan Zhang et al carried out the study on Measurements of laminar burning velocities and flame stability analysis for dissociated methanol–air–diluents mixtures at elevated temperatures and pressures. In this the laminar burning velocities and Markstein lengths for the dissociated methanol–air–diluent mixtures were measured at different equivalence ratios, initial temperatures and pressures. The influences of these parameters on the laminar burning velocity and Markstein length were analyzed. It was found that the peak laminar burning velocity occurs at equivalence ratio of 1.8. The Markstein length decreases with an increase in initial temperature and initial pressure. Measurements of laminar burning velocities and flame Stability analyses are conducted using the outwardly spherical laminar premixed flame for DM–air and DM–air–diluents mixtures. The laminar burning velocity and Markstein length at different equivalence ratios, initial temperatures, initial Pressures and N_2/CO_2 dilution ratios are obtained.Erjiang Hu et al carried out the numerical study on laminar burning velocity and NO formation of the premixed methane–hydrogen–air flames.
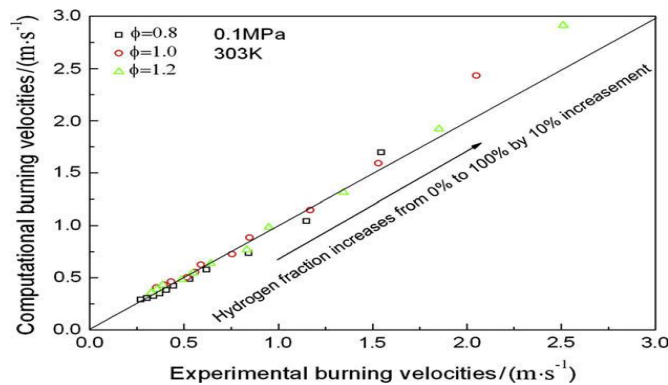
It was found that the unstretched laminar burning velocity is increased with the increase of equivalence ratio and it decreases as the mixtures become fuel-rich.
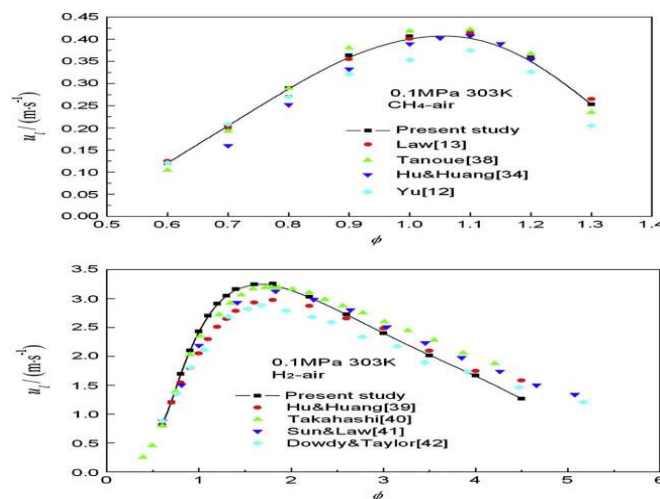
Peak value of unstretched laminar burning velocity of methane–air mixture is presented at the equivalence ratio of 1.1 and that of hydrogen–air mixture is presented at equivalence ratio of 1.8.Methane-dominated combustion is presented when hydrogen fraction is less than 40%, where laminar burning velocity is slightly increased with the increase of hydrogen addition. When hydrogen fraction is larger than 40%, laminar burning velocity is exponentially increased with the increase of hydrogen fraction. With the increase of hydrogen fraction, the overall activation energy of methane–hydrogen mixture is decreased, and the inner layer temperature and Zeldovich number are also decreased. All these factors contribute to the enhancement of combustion as hydrogen is added.

S. Jerzembeck et al carried out the research on Spherical flames of *n*-heptane, iso-octane, PRF 87 and gasoline/air mixtures. These are experimentally investigated to determine laminar burning velocities and Markstein lengths under engine-relevant conditions by using the constant volume bomb method. Data are obtained for an initial temperature of 373 K, equivalence ratios varying from $\phi = 0.7$ to $\phi = 1.2$, and initial pressures from 10 to 25 bar. To track the flame front in the vessel a dark field He–Ne laser Schlieren measurement technique and digital image processing were used. The laminar burning velocities are obtained through a linear extrapolation to zero stretch. The experimentally determined Markstein numbers are compared to theoretical predictions. A reduced chemical kinetic mechanism for *n*-heptane and iso-octane was derived from the Lawrence Livermore comprehensive mechanisms.S.P. Marshall et al are done research on the laminar burning velocity measurements of liquid fuels at elevated pressures and temperatures with combustion residuals. It was found that the laminar burning velocity is a fundamental property of a fuel that affects many aspects of its combustion behavior. Experimental values are required to validate kinetic simulations, and also to provide input for models of flashback, minimum ignition energy and turbulent combustion. A constant volume vessel (rated at 3.4 MPa) in conjunction with a multi-zone model was used to calculate burning velocity from pressure and schlieren data, allowing the user to select data uncorrupted by heat transfer or cellularity. n-Heptane, iso-octane, toluene, ethylbenzene and ethanol were tested over a wide range of initial pressures (50, 100, 200 and 400 kPa), temperatures (310, 380 and 450 K) and equivalence ratios (0.7–1.4), along with tests using combustion residuals at mole fractions of up to 0.3.

Shuang-Feng Wanga et al carried out the study on Laminar burning velocities and Markstein lengths of premixed methane/air flames near the lean flammability limit in microgravity. In this literature the researchers studied the effects of flame stretch on the laminar burning velocities of near-limit fuel-lean methane/air flames have been studied experimentally using a microgravity environment to minimize the complications of buoyancy. Outwardly propagating spherical flames were employed to assess the sensitivities of the laminar burning velocity to flame stretch, represented by Markstein lengths, and the fundamental laminar burning velocities of unstretched flames. Resulting data were reported for methane/air mixtures at ambient temperature and pressure, over the specific range of equivalence ratio that extended from 0.512 (the microgravity flammability limit found in the combustion chamber) to 0.601. Furthermore, the burning velocities predicted by three chemical reaction mechanisms. Additional results of the present investigation were derived for the overall activation energy and corresponding Zeldovich numbers, and the variation of the global flame Lewis numbers with equivalence ratio. The implications of these results were discussed.

II. DISCUSSION

Various literatures on evaluating the laminar burning velocity has been reviewed briefly and it is noticed that laminar burning velocities for hydrocarbon-air mixture are about 0.4 m/s and under turbulent flow conditions are unlikely to be much above 1.5 m/s. The Bunsen burner will accept flow velocities up to five times the burning velocity so that the maximum flow velocity which can be considered with this design of burner must be about 2 m/s. With rising oil prices and global warming being a dominant environmental issue, it seems that the use of alternative fuels in future is inevitable. These leading goals for both energy security and clean air project have resulted in heightened interests in the worldwide utilizations of alternative fuels in burners and engines. One of the most important parameters for any fuel is the laminar burning velocity. This property forms an important input parameter for models of turbulentCombustion and ignition limits. It is also important in engine simulations, which directly affects power output and efficiency. The experimental procedure used in this work follows closely that of Bradley et al, Metghalchi and Keck, Law et al, and Gülder, these groups also investigated spherical expanding flames at high pressures to determine laminar burning velocities of premixed mixtures in a preheated closed vessel with optical access. In these works, data measured over a similar range of stretch rates were extrapolated to zero stretch value following the approach. Lewis, streholow, Vagelopolous among others has found that the relationship between equivalence ratio and burning velocity is in the form of a bell curve approximately care necessary. The first step is entered on Ø=1. In order to calculate the equivalency ratio ϕ. For this experiment a series of calculations are necessary.

$$\phi = (fuel/air)_{actual} / (fuel/air)_{stoic}$$

A significant characteristic that affects burning velocity is the degree of turbulence in the flame. Ideally, the flame being measured should be laminar. A laminar flame has parallel flow lines, and therefore results in a uniform, steady flame. The Reynolds number (Re) is used to determine the state of the flow for a given apparatus. The equation for the Reynolds number is

$$R_e = \rho v d / \mu$$

Most importantly, for Re<2300 the flow is laminar, and for Re > 3200 it is typically turbulent. A laminar flame is precise and sharply defined which is necessary to accurately determine burning velocity. ρ is density of mixture, v is the average velocity, d is diameter of burner tube, μ is dynamic viscosity.
The flame speed is determined by,

$$S_b = dr_b / dt$$

Where, S_b is the flame speed with respect to the burned gas and r_b is the flame front displacement. Therefore, values of the Schlieren radius r_{sch} determined by image tracking of Schlieren cinematography are close to r_b. Flame images were analyzed with an image processing code specifically developed for the experimental configuration to track flame front radii over time. The burned propagating flame speed S_b was determined from first-order least squares fits through four radii adjacent to each point under consideration. Contrast levels were set to define consistently flame fronts at all times.
The unburned unstretched flame speed (S_u) was obtained from continuity,

$$S_b \, \rho_b = S_u \, \rho_u$$

The values ρ_b and ρ_u are the burned and unburned densities of the mixture and were computed with the one dimensional flame code Flame Master.

III.   FUTURE SCOPE
The experimental setup for this project had several flaws that are significant sources for the margin of error in the data.Particular focus should be on a redesign of the apparatus Possible improvements should include a section of greater vertical distance prior to the slot itself to attempt to induce a flow that is more laminar, and generally less twists and turns in the tubing and adapters. For the Bunsen burner apparatus the most significant improvement would be to add a second tube to the apparatus. The tube would surround the existing burner tube, and it would provide a flow of an inert gas, such as nitrogen. This flow would stop the leakage that occurs in a horizontal direction under atmospheric conditions. This is ideal to provide an exact cone shaped flame, which will in turn provide a more accurate measurement for surface area and finally burning velocity. Burning velocity is depends on the size of the burner tube, a detailed analysis of the effect of tube diameter on burning velocity measurements is part of a future study related to this work.

IV.   CONCLUSION
The laminar burning velocity is one of the fundamental properties of a reacting premixed mixture and its reliable data are constantly needed for combustion applications. So far, several techniques for measuring the laminar burning velocity have been used, and for a wide range of temperature, pressure, and fuel rather accurate measurements have been obtained by employing flat or curved flames in stagnation flow propagating spherical flames in combustion vessel or flat flames stabilized on burner. With all those measurement techniques proper care could be taken to remove the effect of flame stretch either during experimentation or through further data processing. Measurements were based on motion picture schlieren photographs of outwardly propagating spherical flames. A heated spherical combustion vessel has been used with systems for fuel injection, ignition, experiment control, data acquisition and high speed schlieren photography. This study focuses on the effects of initial temperature and fuel/air equivalence ratio on the laminar burning velocities of gases and liquids.

REFERENCES
Journal Papers:
[1]   S.Y. Liao, D.M. Jiang, Z.H. Huang, K. Zeng, Q. Cheng, "Determination of the laminar burning velocities for mixtures of ethanol and air at elevated temperatures",Applied Thermal Engineering,vol.27, pp 374-380,2007.
[2]   Xuan Zhang, Zuohua Huang, Zhiyuan Zhang, Jianjun Zheng, Wu Yu, Deming Jiang, "Measurements of laminar burning velocities and flame stability analysis for dissociated methanol–air–diluents mixtures at elevated temperatures and pressures", International journal of Hydrogen Energy vol.34, pp 4862-4875,2009.
[3]   Erjiang Hu, Zuohua Huang, Jianjun Zheng, Qianqian Li, Jiajia He, " Numerical study on laminar burning velocity and NO formation of premixedmethane–hydrogen–airflames", International journal of Hydrogen Energy, vol. 34, pp 6545-6557,2009.

[4]   S. Jerzembeck , N. Peters, P. Pepiot-Desjardins , H. Pitsch, "Laminar burning velocities at high pressure for primary reference fuels and gasoline: Experimental and numerical investigation", Combustion and Flam*e*, vol. 156, pp   292-301,2009.

[5]   S.P. Marshall , S. Taylor , C.R. Stone, T.J. Davies , R.F. Cracknell, "Laminar burning velocity measurements of liquid fuels at elevated pressures and temperatures with combustion residuals", Combustion and Flame vol. 158, pp   1920-1932,2011.

[6]   Shuang-Feng Wanga,, Hai Zhang b, Jozef Jarosinski c, Andrzej Gorczakowski c, Jerzy Podfilipski, "Laminar burning velocities and Markstein lengths of premixed methane/air flames near the lean flammability limit in microgravity", Combustion and Flame vol 157, pp 667-675,2010.

[7]   Ganeshan v.,"internal combustion engine " third edition Published By the Tata McGraw-Hill Publishing Company Limited, 7 West Patel Nagar, New Delhi 110008. pp 372-375.

[8]   Esam M. Mohamed, **"**Burning velocities & flam temperatures of ethanol & butanol-air mixtures". Technical Institute / Babylon 2010.

AUTHOR'S PROFILE

V. J. Katre is a student of final semester of P.G course in Engineering (Heat Power Engineering) of KITS at Ramtek, Nagpur University, India.

S. K. Bhele is an Associate Professor at KITS, Ramtek in the Mechanical Engineering Department. He has 15 years teaching experience. He has published 03 technical papers in International Conference and 05 technical papers in National conferences and is a Life Member of ISTE, SAE, and Associate Member of Institute Engineers. Pursuing PhD from VNIT Nagpur, India.

Dynamical Properties of Carrier Wave Propagating In a Viscous Fluid

[1,]Edison A. Enaibe , and [2,]Osafile E. Omosede

[1,2,]*Department of Physics,Federal University of Petroleum Resources, P.M.B. 1220, Effurun, Delta State, Nigeria.*

ABSTRACT

In this work, we superposed a 'parasitic wave' on a 'host wave' and the behavior of the constituted carrier wave is studied as it flows through an elastic uniform pipe containing a viscous fluid. The dynamical characteristics of the carrier wave as it propagate through the confined space containing the fluid is effectively studied with the use of simple differentiation technique. It is shown in this work that some wave behave parasitically when they interfere with another one, provided they are basically out of phase. The incoherent nature of their source causes the carrier wave to attenuate to zero after a specified time. The phase velocity, the oscillating radial velocity and other basic physical properties of the constituted carrier wave first increases in value before they steadily attenuate to zero. The study provides a means of using the known parameter values of a given wave to determine the basic parameters of another interfering wave which were initially not known. It is established in this study that the process of attenuation of the basic properties that constitute a physical system is not instantaneous but gradual. Since the mechanics of the resident'host wave' would be posing a serious resistance in order to annul the destructive effect of the parasitic interfering wave.The spectrum of the decay process show exemplary behaviour at certain values of the raising multiplier. This behaviour is caused by the high attraction or constructive interference of the combined effects of the 'host wave' and the 'parasitic wave'.

KEY WORDS:*constituted carrier wave, 'host wave', 'parasitic wave', raising multiplier, characteristic angular velocity, group angular velocity.*

I. INTRODUCTION

If a wave is to travel through a medium such as water, air, steel, or a stretched string, it must cause the particles of that medium to oscillate as it passes [1]. For that to happen, the medium must possess both mass (so that there can be kinetic energy) and elasticity (so that there can be potential energy). Thus, the medium's mass and elasticity property determines how fast the wave can travel in the medium.

The principle of superposition of wave states that if any medium is disturbed simultaneously by a number of disturbances, then the instantaneous displacement will be given by the vector sum of the disturbance which would have been produced by the individual waves separately. Superposition helps in the handling of complicated wave motions. It is applicable to electromagnetic waves and elastic waves in a deformed medium provided Hooke's law is obeyed   Interference effect that occurs when two or more waves overlap or intersect is a common phenomenon in physical wave mechanics. When waves interfere with each other, the amplitude of the resulting wave depends on the frequencies, relative phases and amplitudes of the interfering waves. The resultant amplitude can have any value between the differences and sum of the individual waves [2]. If the resultant amplitude comes out smaller than the larger of the amplitude of the interfering waves, we say the superposition is destructive; if the resultant amplitude comes out larger than both we say the superposition is constructive. The interference of one wave say 'parasitic wave' y_1 on another one say 'host wave' y_2 could cause the resident 'host wave' to undergo dampingto zero if they are out of phase. The damping process of y_2 can be gradual, over-damped or critically damped depending on the rate in which the amplitude of the host wave is brought to zero. However, the general understanding is that the combination of y_1 and y_2 would first yield a third stage called the resultant wave say y, before the process of damping sets in. In this work, we refer to the resultant wave as the constituted carrier wave or simply carrier wave.

A 'parasitic wave' as the name implies, has the ability of destroying and transforming the intrinsic constituents of the 'host wave' to its form after a sufficiently long time. It contains an inbuilt raising multiplier λ which is capable of increasing the intrinsic parameters of the 'parasitic wave' to become equal to those of the 'host wave'. Ultimately, once this equality is achieved, then all the active components of the 'host wave' would have been completely eroded and the constituted carrier wave ceases to exist. A carrier wave in this wise, is a corrupt wave function which certainly describes the activity and performance of most physical systems. Thus, the reliability and the life span of most physical systems are determined by the reluctance and willingness of the active components of the resident 'host wave' to the destructive influence of the 'parasitic wave'. Any actively defined physical system carries along with it an inbuilt attenuating factor such that even in the absence of any external influence the system will eventually come to rest after a specified time. This accounts for the non-permanent nature of all physically active matter.

If the wave function of any given active system is known, then its characteristics can be predicted and altered by means of anti-vibratory component. The activity and performance of any active system can be slowed down to zero-point 'death' by means of three factors: (i) Internal factor (ii) External factor, and (iii) Accidental Factor. The internal factor is a normal decay process. This factor is caused by aging and local defects in the constituent mechanism of the physical system. This shows that every physically active system must eventually come to rest or cease to exist after some time even in the absence of any external attenuating influence. The internal factor is always a gradual process and hence the attenuating wave function of the physical system is said to be under-damped.

The external factor is a destructive interference process. This is usually a consequence of the encounter of one existing well behaved active wave function with another. The resultant attenuating wave function under this circumstance is said to be under-damped, over-damped or critically-damped, depending on how fast the intrinsic constituent characteristics of the wave function decays to zero.The accidental factor leads to a sudden breakdown and restoration of the wave function of the physical system to a zero-point. In this case, all the active intrinsic parameters of the physical system are instantaneously brought to rest and the attenuation process under this condition is said to be critically-damped.

The initial characteristics of a given wave with a definite origin or source can best be determined by the use of a sine wave function. However, for the deductive determination of the initial behaviour of a wave whose origin is not certain, the cosine wave function can best be effectively utilized. Generally, we can use the available information of the physical parameters of a wave at any given position and time $y(x + \delta x, t + \delta t)$ to predict the nature of its source andthe initial characteristics when the position and time was $y(x,t)$.The reader should permit the lack of adequate references since no one has ever worked in this area before now.

The organization of this paper is as follows. In section 1, we discuss the nature of wave and interference. In section 2, we show the mathematical theory of superposition of two incoherent waves. The results emanating from this study is shown in section 3. The discussion of the results of our study is presented in section 4. Conclusion and suggestions for further work is discussed in section 5. The paper is finally brought to an end by a few lists of references.

II.   RESEARCH METHODOLOGY

In this work, we superposed a 'parasitic wave' with inbuilt raising multiplier λ on a 'host wave' and the combined effect of the waves is allowed to flow through a narrow pipe containing a viscous fluid. The attenuation mechanism of the carrier wave which is the result of the superposition is thus studied by means of simple differentiation technique.

2.1.Mathematical theory of superposition of waves

Let us consider two incoherent waves defined by the non - stationary displacement vectors

$$y_1 = a\,\beta \cos(\vec{k}\beta.\vec{r} - n\beta t - \varepsilon\,\beta) \quad (2.1)$$

$$y_2 = b\lambda \cos(\vec{k}'\lambda.\vec{r} - n'\lambda t - \varepsilon'\lambda) \quad (2.2)$$

where all the symbols retain their usual meanings. In this study, (2.1) is regarded as the 'host wave' whose propagation depends on the inbuilt raising multiplier $\beta(= 0 \ldots,1)$. While (2.2) represents a 'parasitic wave'

with an inbuilt raising multiplier $\lambda(=0,1,2,\cdots,\lambda_{\max})$. The inbuilt multipliers are both dimensionless and as the name implies, they are capable of gradually raising the basic intrinsic parameters of both waves respectively with time. Let us superpose (2.2) on (2.1), with the hope to realize a common wave function, then

$$y = y_1 + y_2 = a\beta\cos(\vec{k}\beta\,\vec{r} - n\beta\,t - \varepsilon\,\beta) + b\lambda\cos(\vec{k}'\lambda\,\vec{r} - n'\lambda\,t - \varepsilon'\lambda) \qquad (2.3)$$

Suppose, we assume that for a very small parameter ζ, the below equation holds,

$$n'\lambda = \zeta + n\beta \quad (2.4)$$

$$y = a\beta\cos(\vec{k}\,\beta\,.\vec{r} - n\,\beta\,t - \varepsilon\beta) + b\lambda\cos(\vec{k}'\lambda\,.\vec{r} - n\,\beta\,t - \zeta\,t - \varepsilon'\lambda) \qquad (2.5)$$

Again in (2.5), we let

$$\varepsilon'_1 = \zeta\,t + \varepsilon'\lambda \qquad (2.6)$$

$$y = a\beta\cos(\vec{k}\,\beta\,.\vec{r} - n\,\beta\,t - \varepsilon\beta) + b\lambda\cos(k'\lambda\,.\vec{r} - n\,\beta\,t - \varepsilon'_1) \qquad (2.7)$$

For the purpose of proper grouping we again make the following assumption:

$$\vec{k}\,\beta\,.\vec{r} = \vec{k}'\lambda\,.\vec{r} = \xi \qquad (2.8)$$

$$(\vec{k}\beta - \vec{k}'\lambda).\vec{r} = \xi \qquad (2.9)$$

$$y = a\beta\cos\big((\xi - n\,\beta\,t) - \varepsilon\,\beta\big) + b\lambda\cos\big((\xi - n\beta\,t) - \varepsilon'_1\big) \qquad (2.10)$$

We can now apply the cosine rule for addition of angles to reevaluate each term in (2.10), that is,

$$\cos(A \pm B) = \cos A\cos B \mp \sin A\sin B \quad (2.11)$$

$$y = a\beta\left\{\cos(\xi - n\beta\,t)\cos\beta\,\varepsilon + \sin(\xi - n\beta\,t)\sin\beta\,\varepsilon\right\} +$$
$$b\lambda\left\{\cos(\xi - n\,\beta\,t)\cos\varepsilon'_1 + \sin(\xi - n\,\beta\,t)\sin\varepsilon'_1\right\} \qquad (2.12)$$

$$y = \cos(\xi - n\beta\,t)\left\{a\beta\cos\beta\,\varepsilon + b\lambda\cos\varepsilon'_1\right\} + \sin(\xi - n\beta\,t)\left\{a\beta\sin\beta\varepsilon + b\beta\sin\varepsilon'_1\right\} (2.13)$$

For technicality, let us make the following substitutions so that we can further simplify (2.14).

$$A\cos E = a\beta\cos\beta\varepsilon + b\lambda\cos\varepsilon'_1 \qquad (2.14)$$

$$A\sin E = a\beta\sin\beta\varepsilon + b\lambda\sin\varepsilon'_1 \quad (2.15)$$

$$y = A\left\{\cos(\xi - n\beta\,t)\cos E + \sin(\xi - n\beta\,t)\sin E\right\} \qquad (2.16)$$

$$y = A\cos\left\{\xi - n\beta\,t - E\right\} \quad (2.17)$$

$$y = A\cos\left\{(\vec{k}\beta - \vec{k}'\lambda).\vec{r} - n\beta\,t - E\right\} \quad (2.18)$$

The simultaneous nature of (2.15) and (2.16) would enable us to square though them and add the resulting equations term by term, that is

$$A = \sqrt{a^2\beta^2 + b^2\lambda^2 + 2ab\beta\lambda\cos\big((\beta\varepsilon - \varepsilon'\lambda) + (n\beta - n'\lambda)t\big)} \quad (2.19)$$

$$y = \sqrt{a^2\beta^2 + b^2\lambda^2 + 2a\beta b\lambda\cos\big((\beta\varepsilon - \varepsilon'\lambda) + (n\beta - n'\lambda)t\big)} \times \cos\big((\vec{k}\beta - \vec{k}'\lambda).\vec{r} - n\beta\,t - E\big) \quad (2.20)$$

Upon dividing (2.16) by (2.15), we get that

$$\tan E = \frac{a\beta\sin\beta\,\varepsilon + b\lambda\sin\varepsilon'_1}{a\beta\cos\beta\,\varepsilon + b\lambda\cos\varepsilon'_1} \qquad (2.21)$$

$$E = \tan^{-1}\left(\frac{a\beta\sin\beta\,\varepsilon + b\lambda\sin(\varepsilon'\lambda - (n\beta - n'\lambda)t)}{a\beta\cos\beta\,\varepsilon + b\lambda\cos(\varepsilon'\lambda - (n\beta - n'\lambda)t)}\right) \quad (2.22)$$

Hence (2.26) is the resultant wave equation which describes the superposition of the 'parasitic wave' on the 'host wave'. Equation (2.28) represents a resultant wave equation in which the effects of the constitutive waves are additive in nature. However, suppose the effects of the constitutive waves are subtractive and with the view that the basic parameters of the 'host wave' are constant with time, that is, $\beta = 1$ and leave its variation for future study, then without loss of dimensionality we can recast (2.20) and (2.22) as

$$y = \sqrt{(a^2 - b^2\lambda^2) - 2(a - b\lambda)^2\cos\big((n - n'\lambda)t - (\varepsilon - \varepsilon'\lambda)\big)} \times \cos\big((\vec{k} - \vec{k}'\lambda).\vec{r} - (n - n'\lambda)t - E\big)(2.23)$$

where we have redefined the amplitude an the total phase angle as,

$$A = \sqrt{\left(a^2 - b^2\lambda^2\right) - 2(a - b\lambda)^2 \cos\left((n - n'\lambda)t - (\varepsilon - \varepsilon'\lambda)\right)} \quad (2.24)$$

$$E = \tan^{-1}\left(\frac{a\sin\varepsilon - b\lambda\sin((n - n'\lambda)t - \varepsilon'\lambda)}{a\cos\varepsilon - b\lambda\cos((n - n'\lambda)t - \varepsilon'\lambda)}\right) \quad (2.25)$$

Equation (2.25) is the constitutive carrier wave equation necessary for our study. As the equation stands, it is a 'carrier wave', in which it is only the variation in the intrinsic parameters of the 'parasitic wave' that determines the activityof thephysical system which it describes. Henceforth, we have agreed in this study, that the initial parameters of the 'host wave' are assumed to be constant and also they are initially greater than those of the 'parasitic wave'.

2.1 The calculus of the total phase angle E of the carrier wave function
Let us now determine the variation of the total phase angle E with respect to time t. Thus from (2.25),

$$\frac{dE}{dt} = \left(1 + \left(\frac{a\sin\varepsilon - b\lambda\sin((n - n'\lambda)t - \varepsilon'\lambda)}{a\cos\varepsilon - b\lambda\cos((n - n'\lambda)t - \varepsilon'\lambda)}\right)^2\right)^{-1} \times \frac{d}{dt}\left(\frac{a\sin\varepsilon - b\lambda\sin((n - n'\lambda)t - \varepsilon'\lambda)}{a\cos\varepsilon - b\lambda\cos((n - n'\lambda)t - \varepsilon'\lambda)}\right) \quad (2.26)$$

$$\frac{dE}{dt} = \left\{\frac{(a\cos\varepsilon - b\lambda\cos((n - n'\lambda)t - \varepsilon'\lambda)^2}{(a\cos\varepsilon - b\lambda\cos((n - n'\lambda)t - \varepsilon'\lambda))^2 + (a\sin\varepsilon - b\lambda\sin((n - n'\lambda)t - \varepsilon'\lambda))^2}\right\} \times$$

$$\frac{d}{dt}\left(\frac{a\sin\varepsilon - b\lambda\sin((n - n'\lambda)t - \varepsilon'\lambda)}{a\cos\varepsilon - b\lambda\cos((n - n'\lambda)t - \varepsilon'\lambda)}\right) \quad (2.27)$$

After a lengthy algebra (2.27) simplifies to

$$\frac{dE}{dt} = Z \quad (2.28)$$

where we have introduced a new variable defined by the symbol

$$Z = (n - n'\lambda)\left(\frac{b^2\lambda^2 - ab\lambda\cos\left((\varepsilon + \varepsilon'\lambda) - (n - n'\lambda)t\right)}{a^2 + b^2\lambda^2 - 2ab\lambda\cos\left((\varepsilon + \varepsilon'\lambda) - (n - n'\lambda)t\right)}\right) \quad (2.29)$$

as the characteristic angular velocity of the carrier wave. It has the dimension of *rad./s.*Also the variation of the total phase angle with respect to the wave number is given by

$$\frac{dE}{d(k - k'\lambda)} = t\frac{d}{d(k - k'\lambda)}(n - n'\lambda)\left(\frac{b^2\lambda^2 - ab\lambda\cos\left((\varepsilon + \varepsilon'\lambda) - (n - n'\lambda)t\right)}{a^2 + b^2\lambda^2 - 2ab\lambda\cos\left((\varepsilon + \varepsilon'\lambda) - (n - n'\lambda)t\right)}\right) \quad (2.30)$$

2.2 Evaluation of the group angular velocity (ω_g) of the carrier wave function
The group velocity is a well-defined but different velocity from that of the individual wave themselves. This is also the velocity at which energy is transferred by the wave [3]. When no energy absorption is present, the velocity of energy transport is equal to the group velocity [4]. The carrier wave function is amaximum if the spatial oscillatory phase is equal to 1. As a result

$$\cos\left(r\left((k - k'\lambda)\cos\theta + (k - k'\lambda)\sin\theta\right) - (n - n'\lambda)t - E\right) = 1 \quad (2.31)$$

$$r\left(\cos\theta + \sin\theta\right) - t\frac{d}{d(k - k'\lambda)}(n - n'\lambda) - \frac{dE}{d(k - k'\lambda)} = 0 \quad (2.32)$$

$$r\left(\cos\theta + \sin\theta\right) = t\frac{d}{d(k - k'\lambda)}(n - n'\lambda) + \frac{dE}{d(k - k'\lambda)} \quad (2.33)$$

$$r\left(\cos\theta + \sin\theta\right) = t\frac{d}{d(k - k'\lambda)}(n - n'\lambda) + t\frac{d}{d(k - k'\lambda)}(n - n'\lambda)\left(\frac{b^2\lambda^2 - ab\lambda\cos\left((\varepsilon + \varepsilon'\lambda) - (n - n'\lambda)t\right)}{a^2 + b^2\lambda^2 - 2ab\lambda\cos\left((\varepsilon + \varepsilon'\lambda) - (n - n'\lambda)t\right)}\right)$$

$$(2.34)$$

$$v_g = \frac{r}{t} = \frac{d}{d(k-k'\lambda)} \frac{(n-n'\lambda)}{(\cos\theta + \sin\theta)} \left\{ 1 + \left(\frac{b^2\lambda^2 - ab\lambda \cos\left((\varepsilon + \varepsilon'\lambda) - (n-n'\lambda)t\right)}{a^2 + b^2\lambda^2 - 2ab\lambda \cos\left((\varepsilon + \varepsilon'\lambda) - (n-n'\lambda)t\right)} \right) \right\} \quad (2.35)$$

$$v_g = \frac{r}{t} = \frac{d\omega_g}{d(k-k'\lambda)} \quad (2.36)$$

which is the basic expression for the group angular velocity, where

$$\omega_g = \frac{(n-n'\lambda)}{(\cos\theta + \sin\theta)} \left(\frac{a^2 + 2b^2\lambda^2 - 3ab\lambda \cos\left((\varepsilon + \varepsilon'\lambda) - (n-n'\lambda)t\right)}{a^2 + b^2\lambda^2 - 2ab\lambda \cos\left((\varepsilon + \varepsilon'\lambda) - (n-n'\lambda)t\right)} \right) (2.37)$$

is the group velocity of the carrier wave which has the dimension of *radian/s*. Although, ω_g and Z has the same dimension, but where Z depends on time, ω_g is dependent upon the spatial frequency or wave number(k).

2.3 Evaluation of the phase velocity (v_p)of the carrier wave.

The phase velocity denotes the velocity of a point of fixed phase angle [4]. At any instant of the wave motion the displacements of other points nearby change also and there will be one of these points, at $x + \delta x$ say, where the displacement $y(x + \delta x, t + \delta t)$ is equal to the original displacement $y(x,t)$ at point x. Now from(2.23) the carrier wave is a maximum when the spatial oscillatory phase is equal to one.

$$\cos\left((\vec{k} - \vec{k}'\lambda).\vec{r} - (n-n'\lambda)t - E\right) = 1 \quad (2.38)$$

$$(\vec{k} - \vec{k}'\lambda) = (k-k'\lambda)_x i + (k-k'\lambda)_y j + (k-k'\lambda)_z k \quad (2.39)$$

$$\vec{r} = xi + yj + zk \quad (2.40)$$

If we assume that the motion is constant in the z-direction and the wave vector mode is also the same for both x and y plane, then (2.40) becomes

$$\vec{r} = r\cos\theta\, i + r\sin\theta\, j \quad (2.41)$$

where $\theta = \pi - (\varepsilon - \varepsilon'\lambda)$ is the variable angle between y_1 and y_2, please see appendix for details.Hence

$$\cos\left((k-k'\lambda)r\cos\theta + (k-k'\lambda)r\sin\theta - (n-n'\lambda)t - E\right) = 1 (2.42)$$

$$\left((k-k'\lambda)r\cos\theta + (k-k'\lambda)r\sin\theta - (n-n'\lambda)t - E\right) = 0 (2.43)$$

$$\left((k-k'\lambda)\cos\theta + (k-k'\lambda)\sin\theta\right)d\,r - (n-n'\lambda)d\,t - \frac{dE}{dt}dt = 0 \quad (2.44)$$

$$\left((k-k'\lambda)\cos\theta + (k-k'\lambda)\sin\theta\right)d\,r - (n-n'\lambda)d\,t - Zdt = 0 \quad (2.45)$$

$$\left((k-k'\lambda)\cos\theta + (k-k'\lambda)\sin\theta\right)d\,r = \left((n-n'\lambda) + Z\right)dt \quad (2.46)$$

$$v_p = \frac{d\,r}{dt} = \left(\frac{\left((n-n'\lambda) + Z\right)}{(k-k'\lambda)(\cos\theta + \sin\theta)} \right) (2.47)$$

This has the dimension of *m/s*. Since our argument is equally valid for all values ofr, (2.47) tells us that the whole sinusoidal wave profile move to the left or to the right at a speed v_p.

2.4 Evaluation of the oscillating angularfrequency ($\dot{\theta}$) of the carrier wave.

The variation of the spatial oscillatory phase of the carrier wave with respect to time gives theoscillating frequency ($\dot{\theta}$). Hence, from (2.43)

$$\left((k-k'\lambda)r\cos\theta + (k-k'\lambda)r\sin\theta - (n-n'\lambda)t - E\right) = 0 \quad (2.48)$$

$$(k-k'\lambda)r\left(\frac{d}{dt}(\cos\theta + \sin\theta)\right) - (n-n'\lambda) - \frac{dE}{dt} = 0 \quad (2.49)$$

$$(k-k'\lambda)r\left(\frac{d\cos\theta}{d\theta}\frac{d\theta}{dt} + \frac{d\sin\theta}{d\theta}\frac{d\theta}{dt}\right) - (n-n'\lambda) - Z = 0 (2.50)$$

$$(k-k'\lambda)r\left(\cos\theta\dot{\theta} - \sin\theta\dot{\theta}\right) = \left((n-n'\lambda) + Z\right) (2.51)$$

$$v_\theta = \dot{\theta} = \left(\frac{\left((n - n'\lambda) + Z\right)}{(k - k'\lambda)\, r(\cos\theta - \sin\theta)} \right) \quad (2.52)$$

The unit is per second (s^{-1}). Thus because of the tethered nature of the elastic pipe the carrier wave can only possess oscillating radial velocity and not oscillating angular velocity.

2.5   Evaluation of the radial velocity (v_r) of the carrier wave.

The variation of the spatial oscillatory phase of the carrier wave with respect to time gives the radial velocity (v_r). Hence, from (2.48)

$$(k - k'\lambda) \frac{dr}{dt} (\cos\theta + \sin\theta) - (n - n'\lambda) - \frac{dE}{dt} = 0 \quad (2.53)$$

$$(k - k'\lambda) \dot{r} \left((\cos\theta + \sin\theta)\right) - (n - n'\lambda) - Z = 0 \quad (2.54)$$

$$v_r = \dot{r} = \left(\frac{\left((n - n'\lambda) + Z\right)}{(k - k'\lambda)(\cos\theta + \sin\theta)} \right) \quad (2.55)$$

This has the same unit as the phase velocity which is m/s.

2.6   Evaluation of the velocity of the 'carrier wave'

Let us now evaluate the velocity with which the entire carrier wave function moves with respect to time. This has to do with the product differentiation of the non-stationary amplitude and the spatial oscillatory cosine phase.

$$v = \frac{dy}{dt} = (a - b\lambda)^2 (n - n'\lambda) \sin\left((n - n'\lambda)t - (\varepsilon - \varepsilon'\lambda)\right) \times$$

$$\left((a^2 - b^2\lambda^2) - 2(a - b\lambda)^2 \cos\left((n - n'\lambda)t - (\varepsilon - \varepsilon'\lambda)\right)\right)^{-\frac{1}{2}} \times$$

$$\cos\left((k - k'\lambda)\, r(\cos\theta + \sin\theta) - (n - n'\lambda)t - E\right) +$$

$$\left((a^2 - b^2\lambda^2) - 2(a - b\lambda)^2 \cos\left((n - n'\lambda)t - (\varepsilon - \varepsilon'\lambda)\right)\right)^{\frac{1}{2}} \times$$

$$\left\{(n - n'\lambda) + Z - V_r\,(k - k'\lambda)(\cos\theta + \sin\theta) - V_\theta\,(k - k'\lambda)\, r\,(\cos\theta + \sin\theta)\right\} \times$$

$$\sin\left((k - k'\lambda)\, r(\cos\theta + \sin\theta) - (n - n'\lambda)t - E\right) \quad (2.56)$$

$$v = (a - b\lambda)^2 (n - n'\lambda) \sin\left((n - n'\lambda)t - (\varepsilon - \varepsilon'\lambda)\right) \times$$

$$(a^2 - b^2\lambda^2)^{-\frac{1}{2}} \left(1 - \frac{2(a - b\lambda)^2 \cos\left((n - n'\lambda)t - (\varepsilon - \varepsilon'\lambda)\right)}{(a^2 - b^2\lambda^2)}\right)^{-\frac{1}{2}} \times$$

$$\cos\left((k - k'\lambda)\, r(\cos\theta + \sin\theta) - (n - n'\lambda)t - E\right) +$$

$$(a^2 - b^2\lambda^2)^{\frac{1}{2}} \left(1 - \frac{2(a - b\lambda)^2 \cos\left((n - n'\lambda)t - (\varepsilon - \varepsilon'\lambda)\right)}{(a^2 - b^2\lambda^2)}\right)^{\frac{1}{2}} \times$$

$$\left\{(n - n'\lambda) + Z - V_r\,(k - k'\lambda)(\cos\theta + \sin\theta) - V_\theta\,(k - k'\lambda)\, r\,(\cos\theta + \sin\theta)\right\} \times$$

$$\sin\left((k - k'\lambda)\, r(\cos\theta + \sin\theta) - (n - n'\lambda)t - E\right) \quad (2.57)$$

Upon using Binomial expansion on the fractional terms and stopping at the second term we get

$$v = \frac{(a - b\lambda)^2 (n - n'\lambda)}{(a^2 - b^2\lambda^2)^{\frac{1}{2}}} \sin\left((n - n'\lambda)t - (\varepsilon - \varepsilon'\lambda)\right) \times$$

$$\cos\left((k - k'\lambda)\, r(\cos\theta + \sin\theta) - (n - n'\lambda)t - E\right) +$$

$$\frac{(a - b\lambda)^4 (n - n'\lambda)}{(a^2 - b^2\lambda^2)^{\frac{3}{2}}} \cos\left((n - n'\lambda)t - (\varepsilon - \varepsilon'\lambda)\right) \times \cos\left((k - k'\lambda)\, r(\cos\theta + \sin\theta) - (n - n'\lambda)t - E\right) +$$

$$(a^2 - b^2 \lambda^2)^{\frac{1}{2}} \left\{ (n - n'\lambda) + Z - V_r (k - k'\lambda)(\cos\theta + \sin\theta) - V_\theta (k - k'\lambda) r (\cos\theta + \sin\theta) \right\} \times$$

$$\sin\left((k - k'\lambda) r (\cos\theta + \sin\theta) - (n - n'\lambda) t - E\right) - \frac{(a - b\lambda)^2}{(a^2 - b^2 \lambda^2)^{\frac{1}{2}}} \cos\left((n - n'\lambda) t - (\varepsilon - \varepsilon'\lambda)\right) \times$$

$$\left\{ (n - n'\lambda) + Z - V_r (k - k'\lambda)(\cos\theta + \sin\theta) - V_\theta (k - k'\lambda) r (\cos\theta + \sin\theta) \right\} \times$$

$$\sin\left((k - k'\lambda) r (\cos\theta + \sin\theta) - (n - n'\lambda) t - E\right) (2.58)$$

For the velocity of the constituted carrier wave to be maximum we have to ignore all the oscillating phases, so that

$$v_m = \frac{(a - b\lambda)^4 (n - n'\lambda) + (a^2 - b^2 \lambda^2)^2 (n - n'\lambda)}{(a^2 - b^2 \lambda^2)^{\frac{3}{2}}} \quad (2.59)$$

The unit is m/s.

2.7   Evaluation of the energy attenuation equation

In natural systems, we can rarely find pure wave which propagates free from energy-loss mechanisms. But if these losses are not too serious we can describe the total propagation in time by a given force law $f(t)$. The propagating carrier wave which takes its course from the origin of the elastic pipe is affected by two factors: (i) the damping effect of the mass of the surrounding fluid (ii) the damping effect of the dynamic viscosity of the elastic walls of the pipein response to the wave propagation. Let us consider a carrier wave propagating through an elastic pipe of a given elasticity Q, if the fluid in the pipe has a mass m and viscosity μ, the dissipationof the carrier wave-energy if the fluid is Newtonian, would obey the equation

$$f(t) = \mu \frac{d^2 y}{dt} + Q y^2 (2.60)$$

$$f(t) dt = 2\mu y d y + Q y^2 dt (2.61)$$

For the carrier wave to have a maximum value then the spatial oscillating part is ignored such that

$$y_m^2 = \left(a^2 - b^2\lambda^2\right) - 2\left(a - b\lambda\right)^2 \cos\left((n - n'\lambda) t - (\varepsilon - \varepsilon'\lambda)\right) (2.62)$$

$$f(t) dt = 2\mu y_m dy_m + Q\left(a^2 - b^2\lambda^2 - 2\left(a - b\lambda\right)^2 \cos\left((n - n'\lambda) t - (\varepsilon - \varepsilon'\lambda)\right)\right) dt (2.63)$$

$$\int f(t) dt = 2\mu \int y_m dy_m + Q\int \left((a^2 - b^2\lambda^2) - 2\left(a - b\lambda\right)^2 \cos\left((n - n'\lambda) t - (\varepsilon - \varepsilon'\lambda)\right)\right) dt (2.64)$$

$$impulse = 2\mu y_m^2 + Q\left((a^2 - b^2\lambda^2) t - \frac{2\left(a - b\lambda\right)^2 \sin\left((n - n'\lambda) t - (\varepsilon - \varepsilon'\lambda)\right)}{(n - n'\lambda)}\right) (2.65)$$

$$impulse \times v_m = 2\mu y_m^2 \times v_m + Q \times v_m \left((a^2 - b^2\lambda^2) t - \frac{2\left(a - b\lambda\right)^2 \sin\left((n - n'\lambda) t - (\varepsilon - \varepsilon'\lambda)\right)}{(n - n'\lambda)}\right) (2.66)$$

$$Energy\, E = 2\mu y_m^2 v_m + Q v_m \left(\frac{(a^2 - b^2\lambda^2)(n - n'\lambda) t - 2\left(a - b\lambda\right)^2 \sin\left((n - n'\lambda) t - (\varepsilon - \varepsilon'\lambda)\right)}{(n - n'\lambda)}\right) (2.67)$$

The unit is in Joulesor $kgm^2 s^{-2}$ or Nm. However, in this work we assume the dynamic viscosity μ of the fluid mediumwhere the carrier wave is propagating as $\mu = 0.004\, N\, s\, m^{-2}$ and the elasticity of the wall of the narrow pipe $Q = 1.9048 \times 10^{-6}\, kg\, m^{-1} s^{-2}$.

2.8   Determination of the 'parasitic wave' parameters (b , n' , ε' and k')

Now, suppose we assign arbitrary values for the parameters of the 'host wave' say: $a = 0.00524m$, $n = 0.000882 rad./s$, $\varepsilon = 0.6109 rad.$ (35^0) and $k = 166 rad./m$, then we can precisely determine the basic parameters of the 'parasitic wave' which were initially not known before the interference, using the below method. The gradual deterioration in the physical system under study shows that after a sufficiently long period of time all the active constituents of the resident 'host wave' would have been completely eroded by the destructive influence of the 'parasitic wave', on the basis of this argument the below relation holds.

$$a - b\lambda = 0 \implies 0.00524 = b\lambda$$
$$n - n'\lambda = 0 \implies 0.000882 = n'\lambda$$
$$\varepsilon - \varepsilon'\lambda = 0 \implies 0.6109 = \varepsilon'\lambda \} \text{-------------------- (2.68)}$$
$$k - k'\lambda = 0 \implies 166 = k'\lambda$$

Upon dividing the sets of relations in (2.68) with one another with the view to first eliminate λ we get

$$\implies 5.94104n' = b$$
$$\implies 0.0085775\varepsilon' = b$$
$$\implies 0.000031566k' = b$$
$$\implies 0.00144377\varepsilon' = n' \} \text{--------------------(2.69)}$$
$$\implies 0.0000053135k' = n'$$
$$\implies 0.00368k' = \varepsilon'$$

A more realistic and applicable relation is when: $0.0085775\varepsilon' = 0.000031566k'$, from which based on simple ratio

$$\varepsilon' = 0.0000316 rad.$$
$$k' = 0.00858 rad./m \} \text{-------------------- (2.70)}$$
$$n' = 4.56 \times 10^{-8} rad./s$$

$$b = 2.71 \times 10^{-7} m$$

Any of these values of the 'parasitic wave' shall produce a corresponding approximate value of lambda $\lambda = 19332$ upon substituting them into (2.68). Note that for the interest of uniformity and anticipated complications we are using the minimum value which is 19332. Hence the interval of the multiplier is $0 \le \lambda \le 19332$. Thus, so far, we have determined the basic intrinsic parameters of both the 'host wave' and those of the 'parasitic wave' both contained in the carrier wave.

2.9   Determination of the attenuation constant (η)

Attenuation is a decay process. It brings about a gradual reduction and weakening in the initial strength of the basic parameters of a given physical system. In this study, the parameters are the amplitude (a), phase angle (ε), angular frequency (n) and the spatial frequency (k). The dimension of the attenuation constant (η)is determined by the system under study. However, in this work, attenuation constant is the relative rate of fractional change (FC) in the basic parameters of the carrier wave. There are 4 (four) attenuating parameters present in the carrier wave. Now, if a , n , ε , k represent the initial basic parameters of the 'host wave' that is present in the carrier wave and $a - b\lambda$, $n - n'\lambda$, $\varepsilon - \varepsilon'\lambda$, $k - k'\lambda$ represent the basic parameters of the 'host wave' that survives after a given time. Then, the FC is

$$\sigma = \frac{1}{4} \times \left(\left(\frac{a - b\lambda}{a} \right) + \left(\frac{\varepsilon - \varepsilon'\lambda}{\varepsilon} \right) + \left(\frac{n - n'\lambda}{n} \right) + \left(\frac{k - k'\lambda}{k} \right) \right) \text{ (2.71)}$$

$$\eta = \frac{FC\big|_{\lambda=i} - FC\big|_{\lambda=i+1}}{unit\, time\,(s)} = \frac{\sigma_i - \sigma_{i+1}}{unit\, time\,(s)} \text{   (2.72)}$$

The dimension is *per second (s^{-1})*.   Thus (2.72) gives $\eta = 0.0000517 s^{-1}$ for all values of λ $(i = 0, 1, 2, \ldots, 19332)$.

2.10   Determination of the time (t)

We used the information provided in section 2.9, to compute the various times taken for the carrier wave to attenuate to zero. The maximum time the carrier wave lasted as a function of the raising multiplier λ is also calculated from the attenuation equation shown by (2.72). The reader should note that we have adopted a slowly varying regular interval for the raising multiplier since this would help to delineate clearly the physical parameter space accessible to our model. However, it is clear from the calculation that the different attenuating fractional changes contained in the carrier wave are approximately equal to one another.We can now apply the attenuation time equation given below.

$$\sigma = e^{-(2^{\gamma}\eta t)/\lambda} \qquad (2.73)$$

$$t = -\left(\frac{\lambda}{2^{\gamma}\eta}\right)\ln\sigma \qquad (2.74)$$

where γ is the functional index of any physical system under study and here we assume $\gamma = 3$ .The equation is statistical and not a deterministic law. It gives the expected basic intrinsic parameters of the 'host wave' that survives after time t . Clearly, we used (2.74) to calculate the exact value of the decay time as a function of the raising multiplier.In this work, we used table scientific calculator and Microsoft excel to compute our results. Also the GNUPLOT 4.6 version was used to plot the corresponding graphs.

III.   PRESENTATION OF RESULTS
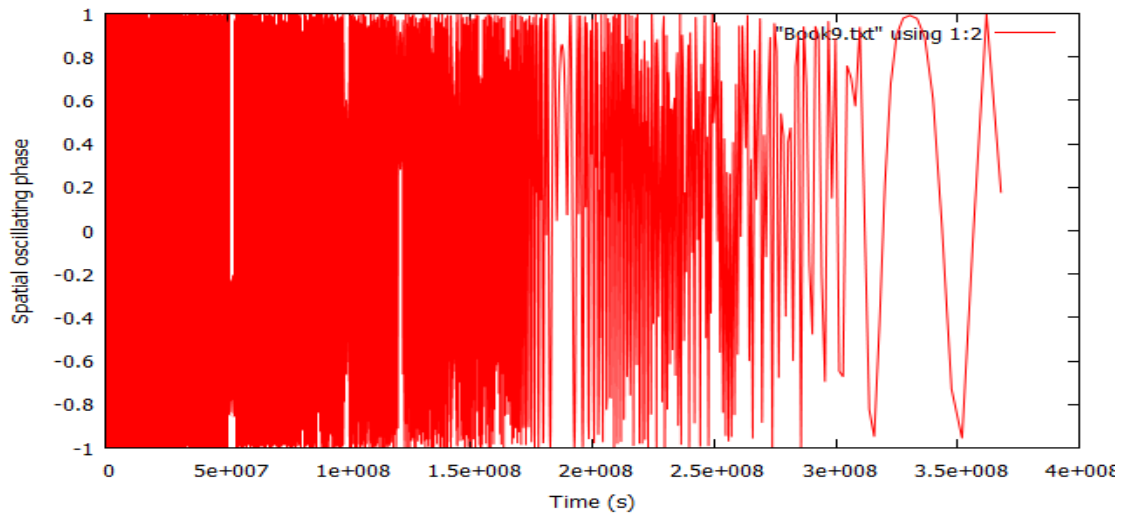
Fig. 3.1. This **r**epresentsthe graph of thespatial oscillating phase of the carrier wave against time t .
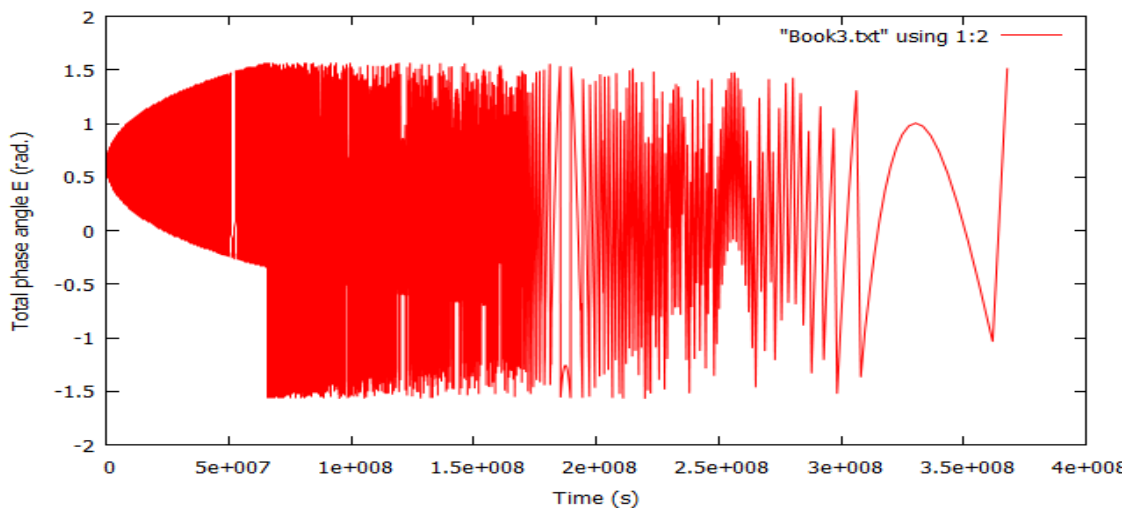
Fig. 3.2. This **r**epresentsthe graph of the total phase angle E of the carrier wave against time t .
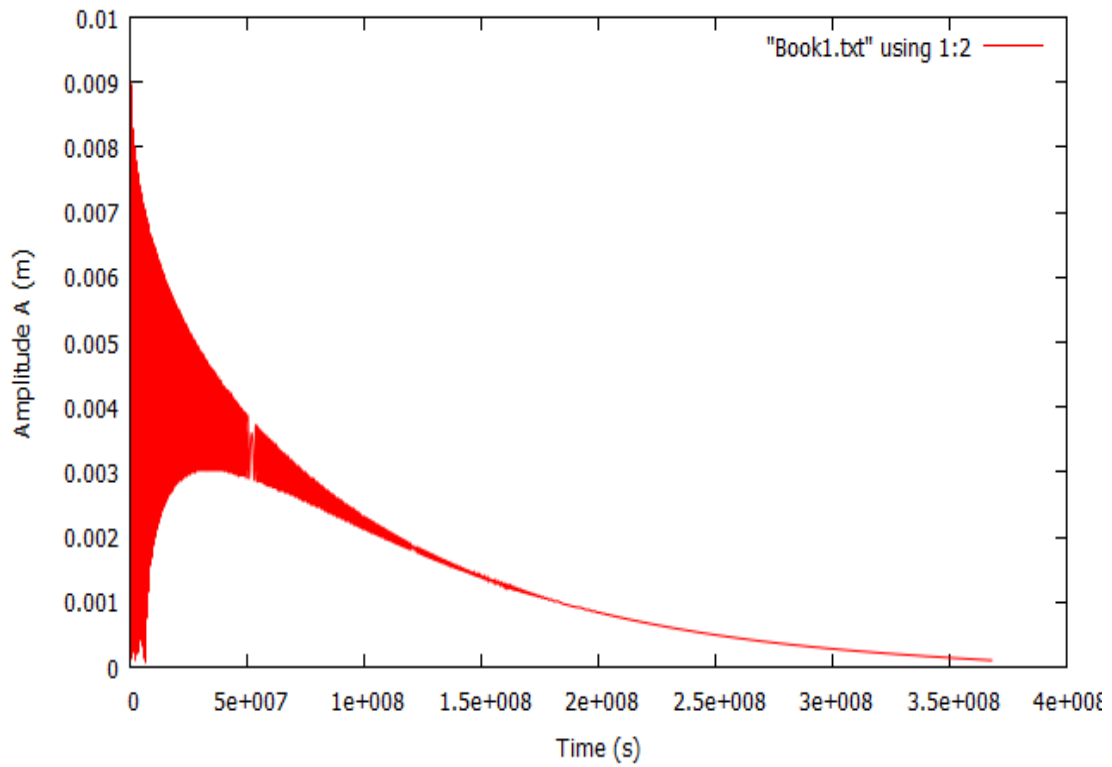
Fig. 3.3. This **r**epresentsthe graph of theamplitude A of the carrier wave against time *t* .
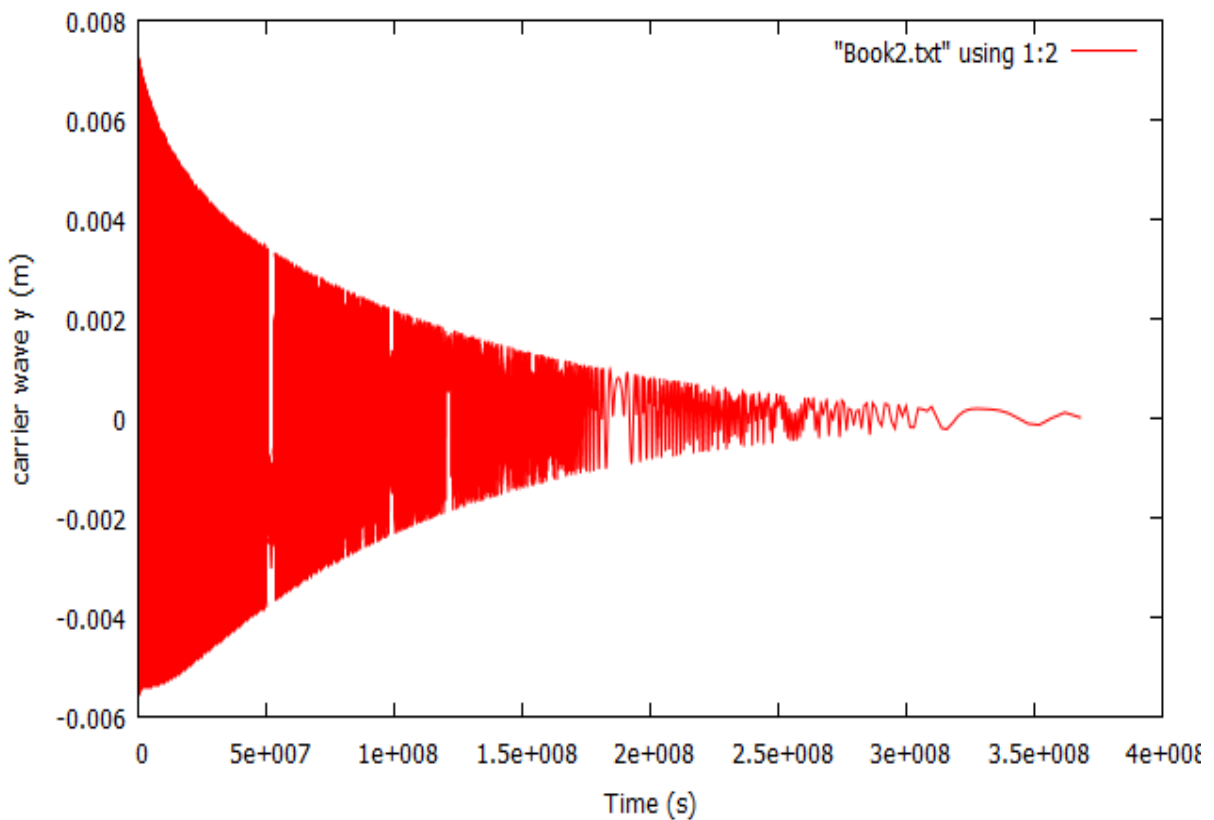
Fig. 3.4. This **r**epresents the graph ofthe carrier wave displacement against time *t* .
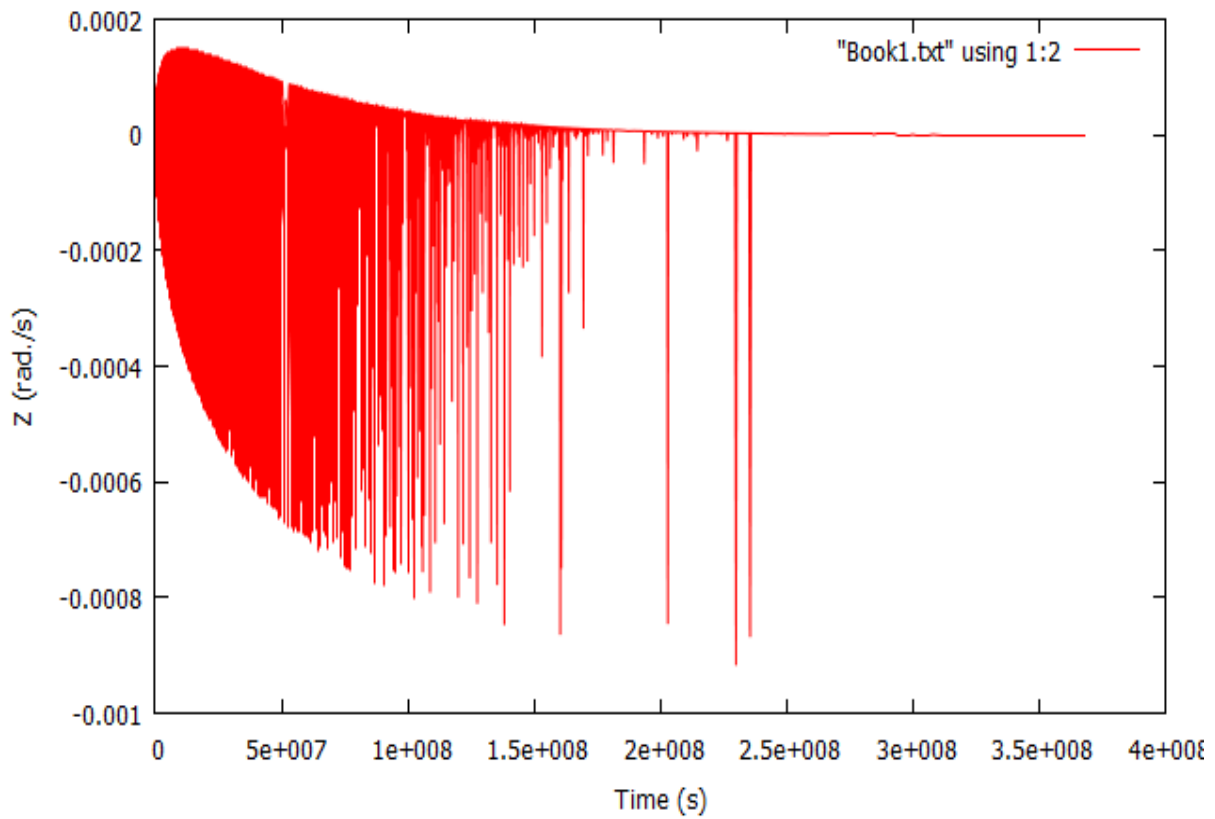
Fig. 3.5. This **r**epresents the graph ofthe characteristic angular frequency Z of the carrier wave against time t .
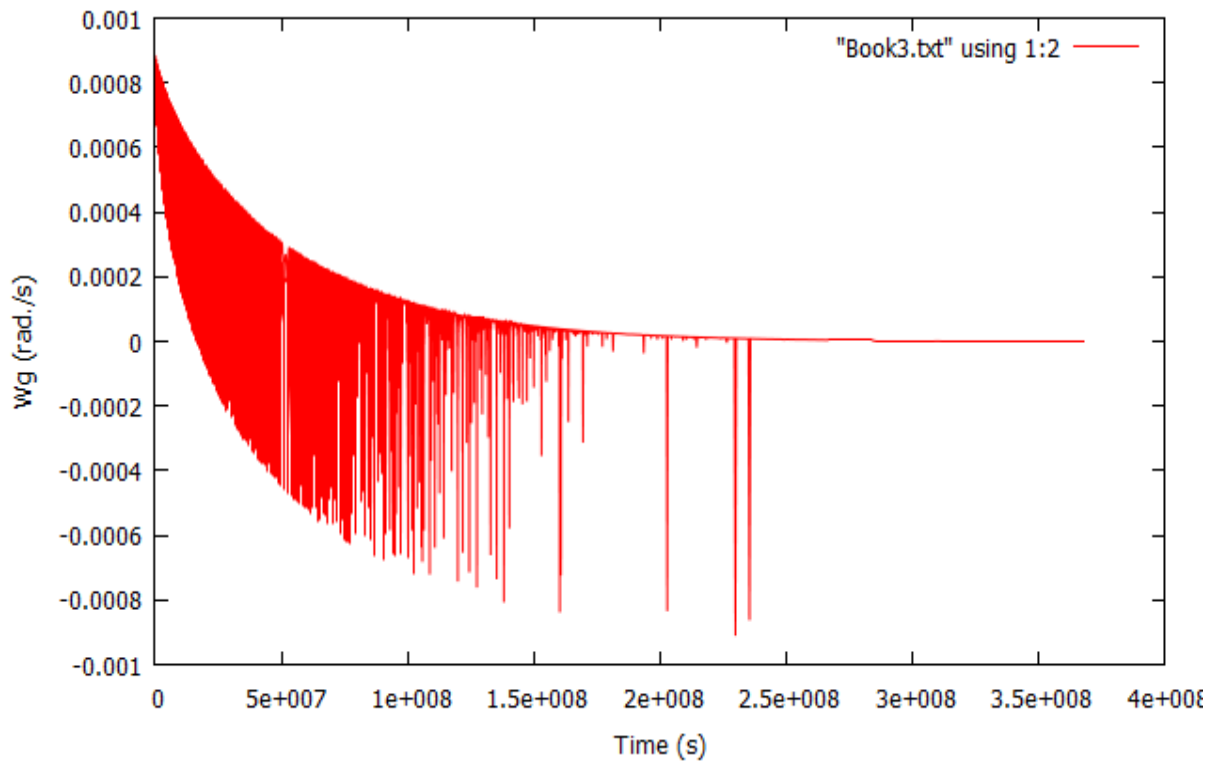
Fig. 3.6. This **r**epresents the graph ofthe group angular velocity Wg of the carrier wave against time t .
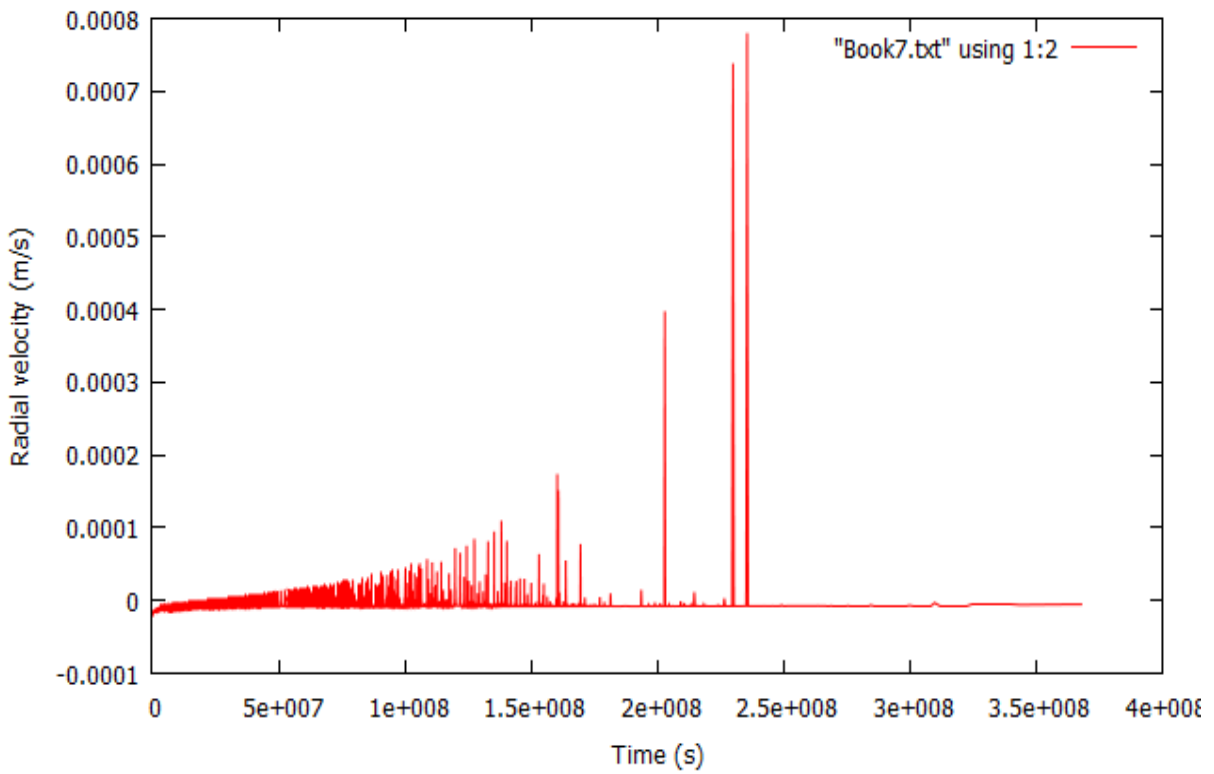
Fig. 3.7. This **r**epresents the graph ofthe radial velocity v_r of the carrier wave against time t .
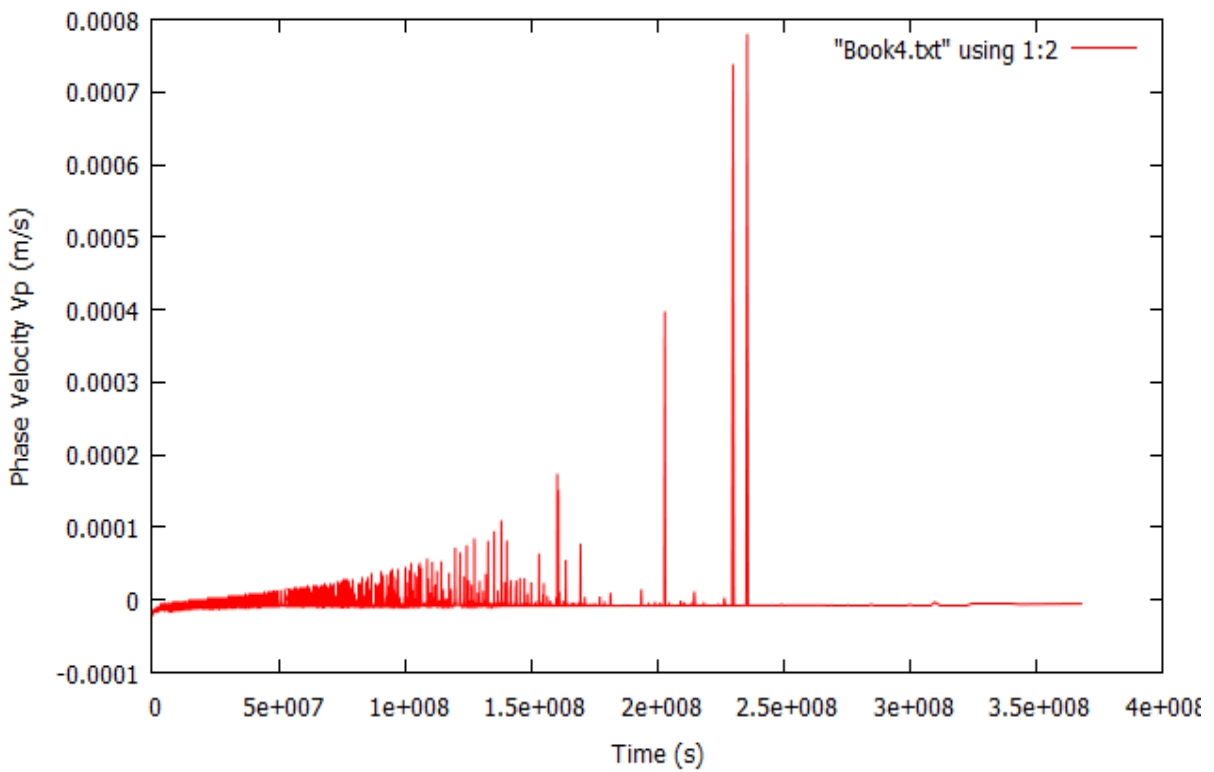
Fig. 3.8. This **r**epresents the graph ofthe phase velocityv_p of the carrier wave against time t .
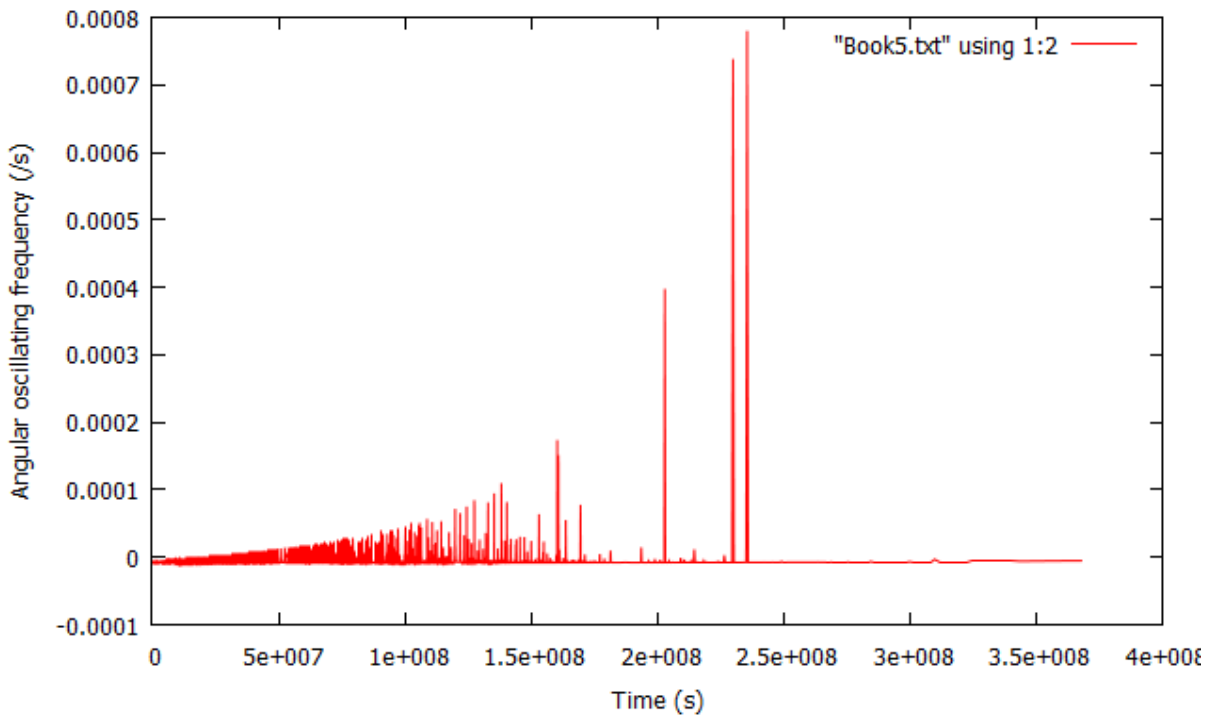
Fig. 3.9. This **r**epresents the graph ofthe angular oscillating frequency ($\dot{\theta}$)of the carrier wave against time t .
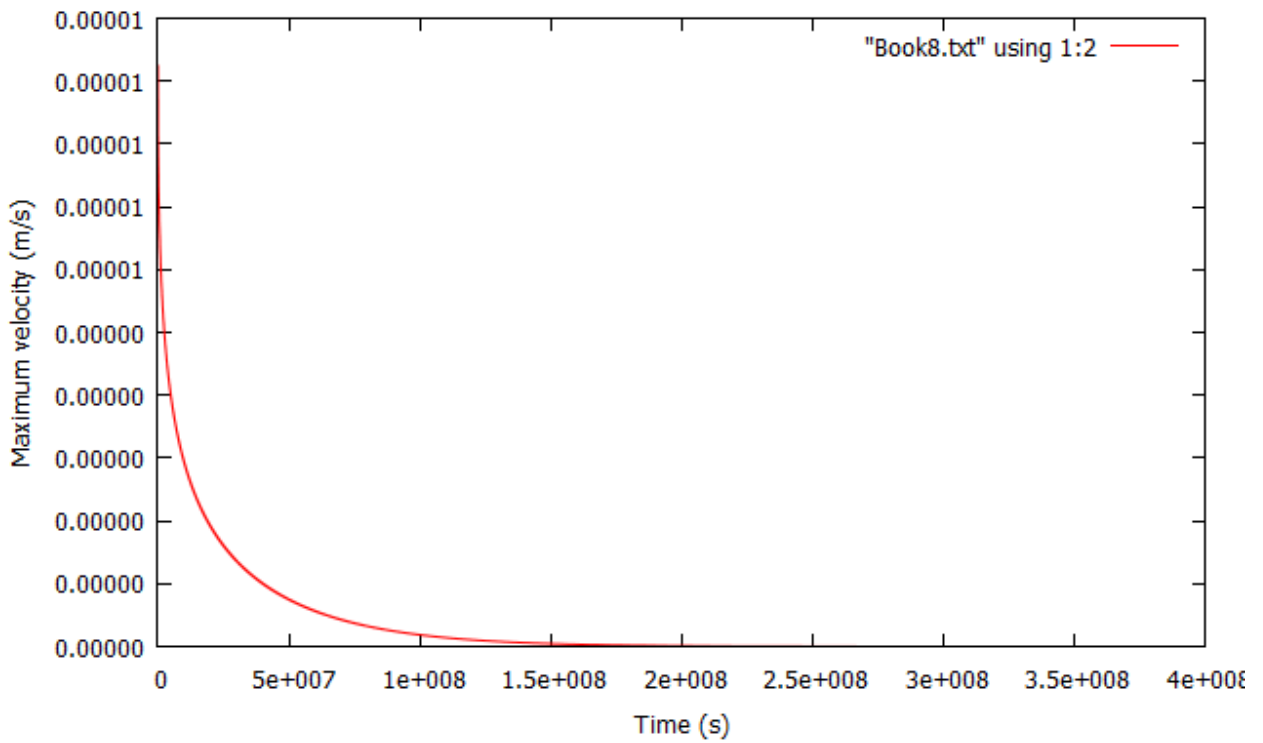
Fig. 3.10. This **r**epresents the graph ofthe maximum velocity $v_{\max}$ of the carrier wave against time t .
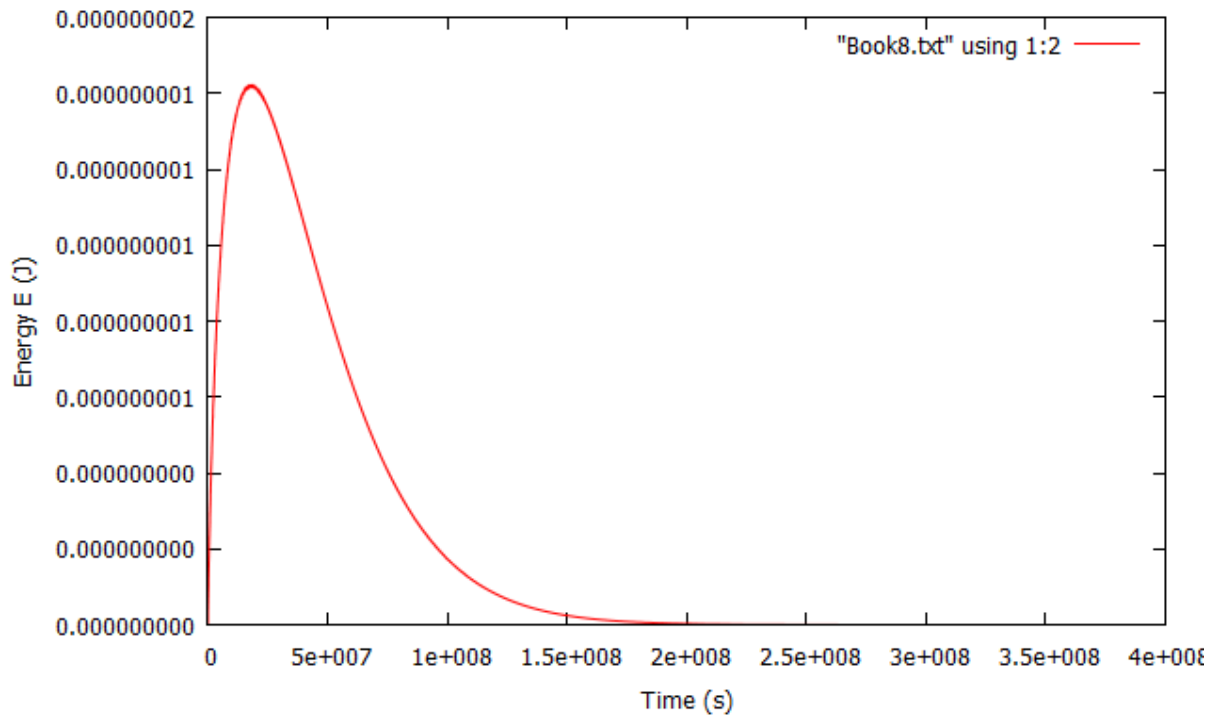
Fig. 3.11. This **r**epresents the graph ofthe attenuated Energy Eof the carrier wave against time t .

IV. DISSCUSION OF RESULTS

The fig.3.1 represents the variation of the spatial oscillating part of the carrier wave with time. It oscillates between the values of ± 1 and converges to zero when the multiplier attainsa maximum value $\lambda = 19332$. Generally, the first significant feature of the spectra given by the figs 3.1, 3.2, 3.3 and 3.4 is the definite singularity at time t = 5 x 10^7s (19 months) or about 2 years. This definite singularity is an indication of the attenuating constituent parameters of the physical system of the resident 'host wave' due to the presence of the interfering 'parasitic wave'. However, the host system further renormalizes to cancel the systems defect and hence the continuous bold spectra.

The second obvious significant feature common to the figures is the depletion in the spectra at t = 1.7 x 10^8 s (65 months) or about 5 years. The interpretation of this depletion is that the interfering 'parasitic wave' is now taking dominant control of the resident host system. Thus the constituent parameters of the 'parasitic wave' are gradually becoming equal to those of the 'host wave'. This results to the decay in the active constituents of the host system. The destructive effect of the interfering wave in the host system is now becoming very intense and difficult to control.Finally, the graphs show that the spectra of the physical system represented by the carrier wave develops failure and hence breakdown if uncontrolled around $t \geq 2.5 \times 10^8 s$ (96 months) or about 8 years. This is marked by the faint and sharp separations in the spectra lines of figs 3.1 - 3.4. The physical system described by the constituted carrier wave goes to zero when $t = 367894196s$ which is about 12 years.

The Fig. 3.2 shows that within the first 2 years (t = 5 x 10^7s) when the multiplicative factor $0 \leq \lambda \leq 14633$, the total phase angle E experiences both positive and negative increase in values. Also the figure reveals a bold spectrum in the total phase angle of the constituted carrier wave in the interval of the multiplicative factor $0 \leq \lambda \leq 18874$and time $0 \leq t \leq 1.7 \times 10^8$s ($0 \leq t \leq 5$years). In this regime the phase angle of the 'host wave' is fluctuating between both positive and negative values thereby undergoing constructive and destructive interactions with that of the 'parasitic wave'. Beyond this interval, that is, when the time $t \geq 1.7 \times 10^8 s$ and the multiplier $18875 \leq \lambda \geq 19332$, the spectrum lines of the total phase angle becomes faint and sharp showing a steady depletion in the total phase angle of the carrier wave.The total phase angle of the constitutive carrier wave has maximum and minimum value of ± 1.5468 rad. However, the spectrum does not finally go to zero rather it diverges even when the multiplier has attained its maximum value $\lambda = 19332$.

In Fig. 3.3, the amplitude of the carrier wave is initially imaginary in the interval $0 \leq \lambda \leq 1866$ and $0 \leq t \leq 122s$ while the amplitude of thecarrier wave is $i\,0.0041863\,lm \leq A \leq i\,0.0040531\,lm$. However, it is the absolute values of the amplitude that we used in the graphical presentation. Subsequently, the amplitude is made up of the imaginary and real part, $A = A_1 + iA_2$ in the interval of the multiplier $0 \leq \lambda \leq 6432$, and the time $0 \leq t \leq 6287092s$ and the amplitude of thecarrier wave is $i\,0.0041863\,lm \leq A \leq i\,0.00018918m$. This shows that the motion is actually two-dimensional (2D). Thus A_1 and A_2 are the components of the amplitude in x and y - directions, and A is tangential to the path of the moving amplitude in the carrier wave. This region of real and imaginary values of the amplitude is an indication of the physical system to guide and renormalizes the system against the effect of the interfering 'parasitic wave'.

There is constant agitation by the intrinsic parameters of the 'host wave' to suppress the destructive influence of the interfering 'parasitic wave' in this region. The intrinsic parameters of the 'host wave' are posing a serious resistance to the destructive tendency of the 'parasitic wave'.The effect of the imaginary decay in the amplitude is unnoticeable or inadequately felt by the physicalsystem in this interval. Although, unnoticeable as it may, but so much imaginary destructive harm would have been done to the intrinsic constituent parameters of the 'host wave'.

Beyond this interval the amplitude of the carrier wave begins to fluctuate with only real values in the interval of the multiplier $6433 \leq \lambda \leq 19233$ and the time 6289274 s $\leq t \leq 241907463$ sor(73 days $\leq t \leq 8$ years)and the amplitude of thecarrier wave is $0.00337988m \leq A \leq 0.00054097m$. In this region the interfering wave is now taking absolute effect on the dynamic mechanical system of the 'host wave'. In other words, counting from the moment the 'parasitic wave' interferes with the 'host wave', there would be absolute indication and manifestation of the 'parasitic wave' after 73 days. The non-consistent attenuating behavior in this interval is a consequence of the fact that the amplitude of the carrier wave do not steadily go to zero, rather it fluctuates. The fluctuation is due to the constructive and destructive interference of both the 'host wave' and the 'parasitic wave'. In the regions where the amplitude of the carrier wave is greater than either of the amplitude of the individual wave, we have constructive interference, otherwise, it is destructive.

Our calculation shows that there is a steady exponential decrease in the values of the amplitude in the interval $242359220s \leq t \leq 367894196s$ (8 years $\leq t \leq 12$ years) when the raising multiplier $19234 \leq \lambda \leq 19332$. This consistent decrease leads to a gradual reduction and weakening in the initial strength of theconstituents of the system of the host. Consequently, the amplitude of the constituted carrier wave consistently attenuates to zero when the raising multiplier $\lambda \geq 19332$ and the time $t \geq 367894196$ s or about 12 years. The amplitude of thecarrier wave thus varies asymptotically between $0.00053612m \leq A \leq 0.00010378m$.

The graph of the carrier wave against time is shown in fig. 3.4. Of course we know that the carrier wave is the product of the amplitude and the spatial oscillating phase. Consequently, the calculated values of the amplitude which is the maximum displacement from some origin are usually greater those of the carrier wave.
As we have said before now, the spectrum of the carrier wave is similar to those of figs, 3.1 and 3.2. The carrier wave experiences a steady damping process and it is consistently attenuated to zero. The spectrum of the carrier wave was initially bold with several regular discontinuities most noticeably in the time interval 1.5 x 10^8 s $\leq t \leq 2$ x 10^8 s (about 5 years). The attenuation of the carrier wave to zero is rapid in the interval when the raising multiplier $19234 \leq \lambda \leq 19332$ and the time 2.5 x 10^8 s $\leq t \leq 4$ x 10^8 s. This of course represents the interval of the predominance of the dynamic constituents of only the interfering 'parasitic wave',while those of the resident 'host wave' are critically undergoing damping.

Thus the phenomenon of AIDS actually occurs in the interval when the raising multiplier $19234 \leq \lambda \leq 19332$ and the time 242359220 s $\leq t \leq 367894196$ s or (8 years $\leq t \leq 12$ years). Consequently, within this interval the physical system under study can no longer annul the destructive effect of the 'parasitic wave'. The carrier wave which describes the coexistence of the 'host wave' and the 'parasitic wave' ceases to exist around 12 years after interference. This is as a result of the fact that all the active constituents of the 'host wave' would have been completely attenuated by the influence of the interfering 'parasitic wave'.

The spectrum of the characteristic angular velocity Z and that of the group angular velocity w_g is represented in figs. 3.5 and 3.6. They are both negatively attractive showing a significant affinity between the resident 'host wave' and the 'parasitic wave'. It is ctear from the fiures that both physical quantities of the carrier wave attenuate to zero after a sufficiently long time. However, the spectrum of the characteristic angular velocity has a wider spectrum than that of the group angular velocity.The spectrum of figs. 3.7, 3.8 and 3.9 are also very similar. The radial velocity, the phase velocity and the oscillating angular frequency of the carrier wave increases positively with increase in time. They all attenuate to zero at $t \geq 2.4 \times 10^8$ s (about 8 years). This exemplary behavior is synonymous with most physical system to produce enhanced efficiency before the system malfunction or complete failure sets in.

The maximum velocity attainby the carrier wave is shown in fig. 3.10. The velocity decreases to zero when the time $t \geq 2 \times 10^8$ (about 6 years). Although, the velocity is zero but there is still a residual velocity that keeps the carrier wave going.

Finally, the energy of the constituted carrier wave as shown in fig. 3.11, first increases beforeit decreases exponentially to zero when the time is $t \geq 2 \times 10^8$ s. The explanation is that the energy attenuation process is not instantaneous and consistent. The lack of consistency is as a result of the constituents of the resident 'host wave' in the carrier wave putting a serious resistance to the destructive influence of the interfering 'parasitic wave'.

V. CONCLUSION

The existence and life span of any physically active system described by the constituted carrier wave is thus determined by the functional index γ. If the index factor issay zero $\gamma = 0$, the carrier wave would have lasted for 2943609286 s (93 years)before it would have attenuated to zero. This study shows that the process of attenuation in most physically active system does not obviously begin immediately. The wave function that defines the activity and performance of most system is guided by some internal factor which enables it to resist any external interfering influence which is destructive in nature. The anomalous behaviour exhibited by the carrier wave at some point during the damping, is due to the resistance pose by the carrier wave in an attempt to annul the destructive effects of the interfering wave. It is evident from this work that when a carrier wave is undergoing attenuation, it does not steadily or consistently come to rest; rather it shows some resistance at some point in time during the damping process, before it finally comes to rest. Consequently, the existence or the life span of any physically active system is determined by the resistance ofits basic intrinsic parameters to the destructive influence of any external factor.

5.1 Suggestions for further work

This study in theory and practice can be extended to investigate wave interference and propagation in three- dimensional (3D) system. The constituted carrier wave we developed in this work can be utilized in the deductive and predictive study of wave attenuation in exploration geophysics and telecommunication engineering. This work can also be extended to investigate energy attenuation in a HIV/AIDS patient.

APPENDIX: Vector representation of the superposition of the 'host wave' and the 'parasiticwave'.
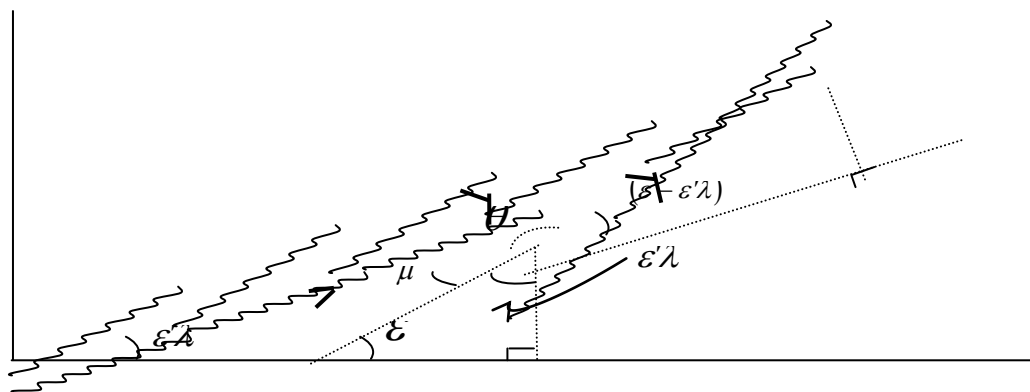
Fig.A1. This represents the resultant wave or the constitutive carrier wave *y* after the superposition of the 'parasitic wave' y_2 on the 'host wave' y_1.

$$\mu + \varepsilon'\lambda + 180^0 - \varepsilon = 180^0 \; ; \mu = \varepsilon - \varepsilon'\lambda \; ; \theta = 180^0 - \left(\varepsilon - \varepsilon'\lambda\right); \theta = \pi - \left(\varepsilon - \varepsilon'\lambda\right)$$

REFERENCES

[1] **David Halliday, Robert Resnick and Jearl Walker**
Fundamentals of Physics, 6th Edition, John Weley and Sons, Inc. New york 2001 : p378.

[2]. **Lain G. Main (1995).** Vibrations and waves in physics
Cambridge University Press, third edition.

[3] **Lipson S.G., Lipson H. and Tannhauser (1996).** Optical physics
Cambridge University Press third edition.

[4] **Brillouin L. (1953).** Wave propagation in periodic structure
Dover, fourth edition New York.

Enhanced Energy Efficient Routing for Self Adapive Geographic Routing In Manet's

[1,]Kalpana M.N , [2,]Mrs.Sridevi K.N   [3,]Dr. Jitendranath Mungara

[1,]*M.Tech,CSE Assoc. Professor*   [2,]*Professor & Dean Dept. of CSE,CMRIT*   [3,]*Dept. of CSE, CMRIT*

ABSTRACT

Geographic routing has been widely used as the most important approach to wireless routing. It has been a big challenge to develop a routing protocol that can meet different application needs and optimize routing paths according to the topology changes in mobile ad hoc networks. In existing system searching of neighbors that are close to the destination was taking lager range of hops which was inefficient due to loss of energy. In this work, we propose an enhanced energy efficient self adaptive geographic routing protocol which reduces the hops count and it is energy efficient compare with existing geographic routing solutions. Our simulations results shows better performances than those with previous approaches..

KEYWORDS: *self adaptive, beacon, wireless communication, Geographic routing, self adaptive, ad hoc network, Greedy, zones.*

I.   INTRODUCTION

A Mobile Ad Hoc Network, also called a MANET, is an autonomous collection of mobile nodes forming a dynamic wireless network. The administration of such a network is decentralized, i.e. each node acts both as host and router and forwards packets for nodes that are not within transmission range of each other. A MANET provides a practical way to rapidly build a decentralized communication network in areas where there is no existing infrastructure or where temporary connectivity is needed.e.g. emergency situations, disaster relief scenarios, and military applications.The topology of a Mobile Ad Hoc Network (MANET) is very dynamic, which makes the design of routing protocols much more challenging than that for a wired network. The conventional MANET routing protocols can be categorized as proactive[1] , reactive[5][6][7] and hybrid[2][3][4] .The proactive protocols maintain the routing information actively, while the reactive ones only create and maintain the routes on demand. The hybrid protocols combine the reactive and proactive approaches. The proactive protocols incur high control overhead when there is no traffic, while for on-demand protocols, the network-range or restricted-range flooding for route discovery and maintenance limits their scalability, and the need of search for an end-to-end path prior to the packet transmission also incurs a large transmission delay.

In recent years, geographic unicast[10][11] and multicast routing[12] have drawn a lot of attentions. Location based routing protocols are the kinds of routing protocols, which use of nodes' location information ,instead of links' information for routing. They are also known as position based routing. In position based routing protocols, it is supposed that the packet source node has position information of itself and its neighbors and packet destination node. In geographic unicast protocols, an intermediate node makes packet forwarding decisions based on its knowledge of the neighbors' positions and the destination's position inserted in the packet header by the source.
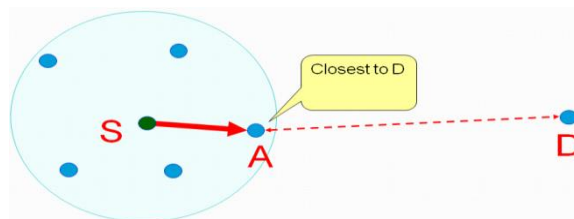
Figure .1. Example of geographic routing.

By default .the packets are transmitted greedily to the neighbor that allows the packet forwarding to make the greatest geographic progress towards the destination.

In existing system schemes[13] we will search route in Network with these conditions:-

In SOGR-HR scheme

- 1-Hop neighbor information.
- 2-Hop neighbor information.
- Until max hops.

In SOGR-GR scheme
- It adopts a reactive beaconing mechanism which is adaptive to the traffic need. Which is more overhead.

The problem with our existing system is searching of hops range is larger due to not finding the neighbors that are close to destination but still it consumes a lot ofenergy.

Figure.2.Problem with existing system

In our work SOGR-EE, the total network is split into many regions. Each Region has a fixed BS. All nodes entering the region send a register message to the BS and BS forwards it to the centralized location service. BS is connected to Centralized location service through a different network (say DSL etc). In the Centralized location service, the region in which the node is present is stored.When any node wants to send data to another node, he can query for the location using the BS in region he stays. BS queries the Centralized location service to provide the region. Source nodes send the data using GPSR routing to the BS in which destination node is residing. From that BS, the message reaches to the destination node according to the SOGR-HR routing mechanism defined in existing solution.

II. RELATEDWOK

In geographic routing, the forwarding decision at each node is based on the locations of the node's one-hop neighbors and location of the packet destination as well. A forwarding nodes therefore needs to maintain these two types of locations. Many works have been proposed to discover and maintain the location of destination. However, the maintenance of one-hop neighbors' location has been often neglected. .Some geographic routing schemes simply assume that a forwarding node knows the location of its neighbors. While others uses periodical beacon broadcasting to exchange neighbors' locations. In the periodic beaconing scheme each node broadcasts a beacon with a fixed beacon interval.In the discussion of geographic routing mechanisms we use the following assumptions:

[1]   Each node knows its geographic location using some localization mechanism [8][9]. Location awareness is essential for many wireless network applications, so it is expected that wireless nodes will be equipped with localization techniques. Several techniques exist for location sensing based on proximity or triangulation using radio signals, acoustic signals, or Infrared. These techniques differ in their localization granularity, deployment complexity, and cost. In general, many localization systems have been proposed in the literature: GPS (Global Positioning System), infrastructure based localization systems and ad-hoc localization systems .
[2]   Each node knows its direct neighbors' locations. This information could be obtained by nodes periodically or on request broadcasting their locations to their neighbors.
[3]   The source knows the destination location.
       A proactive fixed-interval beaconing scheme commonly adopted in existing geographic routing protocols may not only result in a high signaling cost but also outdated local topology knowledge at the forwarding node, which leads to non-optimal routing and forwarding failures. we consider a reference transmissions range to analysis[14] .To explain why the outdated local topology knowledge may lead to non-optimal routing, let us look at the example in Node B just moved into A's transmission range, which is unknown to A before B sends out its next beacon message. Without knowing any neighbors closer to the destination G,
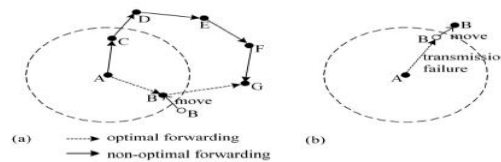
Figure.3.Negative effects of outdated topology information geographic routing.

The resulted path has five hops, while the optimal path between A and G should have only two hops after B bridges the void between A and G. Due to the lack of timely and larger range topology information, the inaccuracy of the local topology knowledge greatly affects the geographic routing performance.In a neighbor's information will be removed if not updated within the timeout interval, which is often set to be multiple beacon intervals. As a result, a node may hold an outdated neighbor information, thus resulting in forwarding failure (e.g., Fig. 3 (b)). This would lead to packet dropping or rerouting [19]. More severely, before detecting the unreachability, the continuous retransmissions at MAC layer reduce the link throughput and fairness, and increase the collisions. This will further increase the delay and energy consumption.
Even though geographic routing has many advantages in existing system schemes.

1. SOGR-HR(SOGR with hybrid reactive mechanism).
2. SOGR-GH(SOGR with Geographic based reactive mechanism).

Both protocols contain an adaptive route optimization component which the position of a next-hop node is estimated before the transmission to avoid position outdate and transmission failure, and the route is optimized according not only to the topology change but also to the actual data traffic requirements.In First scheme the forwarding node forwards the packet to neighbors that are close to the destination with keeping condition hop count H=1 if we are not finding the closest neighbors to the destination with one hop condition then again it searching for Hop count H=2 still if it not finds clostest neighbors it reaches max hops which is more overhead in MANETS. In this scheme we are calculating backoff period for each node to determine whether neighbors nodes closer to destination reply sooner and suppress other's reply..

$$T_{fb}^{N} = \alpha * h * I_{bf} * (1 - \frac{dis_{(F\_D)} - dis_{(N,D)}}{h*} \frac{}{R})$$

where *R* is the reference transmission range of mobile nodes.

In Second scheme (SOGR-GR) it adopts a reactive beaconing Mechanism which is *adaptive* to the traffic need. The periodic beacon is used only when a node overhears data from its neighbors first time. The beaconing is stopped if no data is heard for a predefined period even if it is efficient mechanism cost of beaconing message is more. In Both the scheme the cost of beaconing and route searching is energy consuming. To overcome the drawbacks of both scheme we are using the proposed Energy efficient algorithm (SOGR- EE) which reduces the searching of hops count in network. Our performance studies demonstrate that our algorithm achieve higher delivery ratio, lower control overhead and delay, and energy efficient mechanism in all scenarios tested, with the variations of mobility, traffic, node density and Inaccuracy of destination position. The procedures for finding the next-hop forwarders proposed by existing beaconless schemes may be used with our algorithms and protocols which will help to support more robust and efficient transmissions in various Dynamic conditions.

III.   PROBLEM WITH EXSITING SYSTEM

The problem with both existing system scheme's is due to forwarding node F will try to forward a packet to a neighbor closest to the destination D .In this hop number is restricted to 1 i.e.; h=1.This method has no particular IP address or it has no particular topology if we are not able to find the any neighbors which are more optimum close to destination again it goes for second condition. In this hop number is not restricted, initially hop is set as h=2,it can reach to until maxhops.
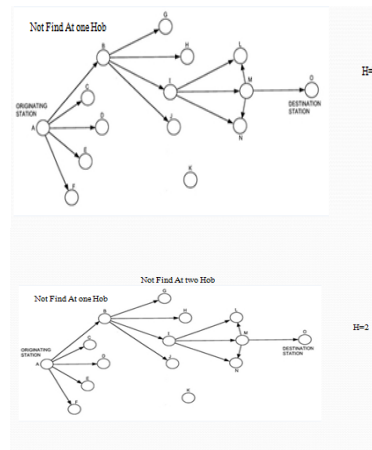
Figure.4.Hops condition

In both protocols the cost of beaconing and route search is energy consuming And Lot of inconsistent routes remain cached in intermediate route which leads of high probability of routing failures. Hop searching range will be larger because of no closer neighbor.

IV.   PROPOSED SYSTEM

The aim of the project work is to design an architectural framework for creating the environment of mobile ad hoc to design the enhanced energy efficient routing for self adaptive geographic routing in MANETs in order to avoid the max hop count with using energy efficient routing. The main objectives of the proposed research work can be briefed as following:

- To design a novel energy efficient routing for Geographic routing (SOGR-EE) in the MANET.
- The number of messages on network (beacons and search messages) are less compared to other messages, so there is less congestion in the network.
- By avoiding beacons [energy reduces due to transmission of beacon packets, receiving of beacon packet] energy consumption is also reduced.
- In a limited zone, the maintenance of routing information is easier.
- All nodes proactively store local routing information, route requests can be more efficiently performed without querying all the network nodes.
- Our proposed routing mechanism helps us to reduce energy consumption due to splitting of entire network into zones .so hop search range will be in smaller range .
- To analyze the throughput with respect to packet delivery ration, energy consumption, Overhead.

In proposed system User configures the system as well as the view the output results using the viewer. User can configure the node movement, location to be present for the Node. User configures the location of the base station. User configures the location service with the expiry timer for the cached entries. Node sends location update to the Base Station, which forwards to the Location service. Two routing engines are present in the system which each Node can use. GPRS routing engine is used to deliver message to the base station GPSR allows nodes to figure out who its closest neighbors are (using beacons) that are also close to the final destination the information is supposed to travel and the SOGR-HR is used to deliver message from base station to the target node. Nodes communicate with each other using wireless channel simulator module. Statistics module collects energy consumed reports from the nodes and delivery reports from the wireless channel simulator and provides to the viewer for presenting to the user.In our work SOGR-EE, the total network is split into many regions. Each Region has a fixed BS. All nodes entering the region send a register message to the BS and BS forwards it to the centralized location service. BS is connected to Centralized location service through a different network (say DSL etc). In the Centralized location service, the region in which the node is present is stored.
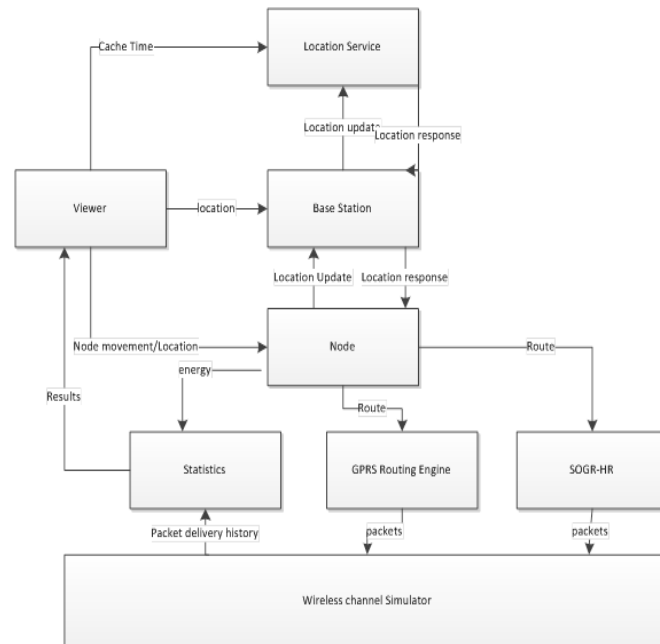
Figure. 5. Proposed Architecture for SOGR-EE

When any node wants to send data to another node, he can query for the location using the BS in region he stays. BS queries the Centralized location service to provide the region. Source nodes send the data using GPSR routing to the BS in which destination node is residing. From that BS, the message reaches to the destination node according to the SOGR-HR routing mechanism defined in existing solution.

4.1.Pseudo code for SOGEE
doRoutingin SOGR-EE(int from , int to)
Input: Source, Destination Output : path
Step 1: Find the current Zone Base Station (B1) of the Source node
Step 2: Find path to base station of the current Zone.
Step 3: Source node query for path to destination node with base station (B1).
Step 4: Base station (B1) gets the destination address (Zone which it is present) from the central station
Step 5: Packet Sent directly to the Base station (B2) of the destination node Zone.
Step 6: B2 sends packet to destination.

V.   PERFORMANCE EVALUATION
In this section, we evaluate the performance of SOGR-HR and SOGR-EE with packet delivery ration, Energy Consumed, Latency. The simulations were run with 100 nodes randomly distributed in an area of 100*m* * 100*m*. We chose a rectangular network area to obtain a longer path.
We study the following metrics:-

Packet delivery ratio :-.the ratio of the number of delivered data packet to the destination. This illustrates the level of delivered data to the destination.

Energy consumption:- Energy consumption is the consumption of energy or power.

Latency :-how much time it takes for a packet of data to get from one designated point to another.
To demonstrate the effectiveness of our algorithms and protocols in supporting robust communications under various conditions, we have performed extensive simulations with packet delivery ratio, energy consumption and latency. In each performance study, only the parameter to evaluate is varied, and the remaining parameters are set to the default values.
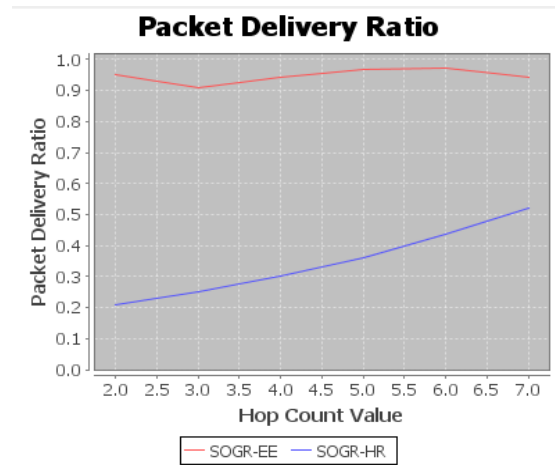
Figure 6.a Packet delivery ratio
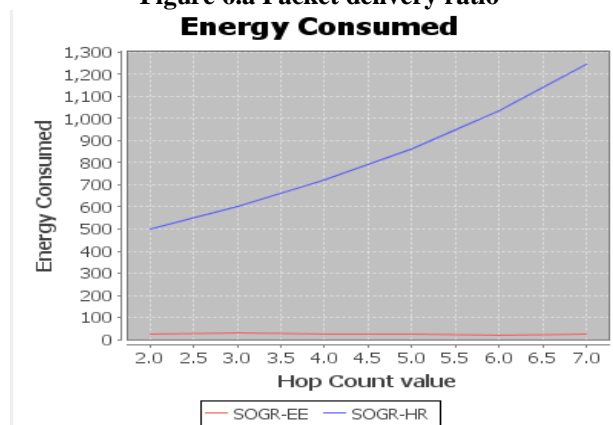
Figure 6.b Energy consumed
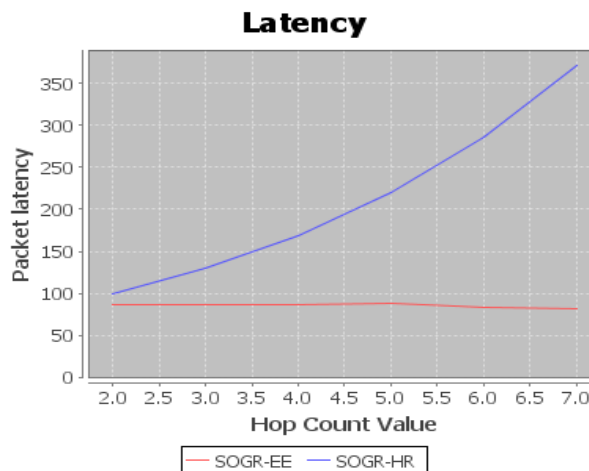
Figure 6.c Latency

In figure 6.(a) the packet delivery ratio of SOGR-HR is reduced as the hop count value increase dramatically but in the case of SOGR-EE with the less hop count value the packet delivery ratio is more and much fig 6.(b)Energy consumption is less in case SOGR-EE even with the increasing hop count value but it is not same in case of SOGR-HR with more hop count value the energy consumption is also more and in fig 6.(c) with more hop count value even with packet latency is more in case of SOGR-HR but in case SOGR-EE with large hop count value the more packet latency is more. The simulation results shows that SOGR-EE is much more efficient than SOGR-HR and it consumes less energy during routing process.

VI.   CONCLUSION N FUTUREWORK

We proposed energy efficient self adaptive geographic routing mechanism which is energy efficient routing in mobile ad networks. We compared our simulations results with existing system SOGR-HG. Our proposed routing mechanism helps us to reduce energy consumption due to splitting of entire network into zones .So hop search range will be in smaller range and it helps us to calculate data deliver ration, latency and energy consumption. In future we can extend the work for multicast communication among different zones for efficient routing.

REFERENCES

[1].   Z. Haas, M. Pearlman and P. Samar. Zone routing protocol (ZRP). Internet draft, draft-ietf-manet-zone- zrp-02.txt, July 2002.
[2].   V. Ramasubramanian, Z. J. Haas and E. G. Sirer. Sharp: A hybrid adaptive routing protocol for mobile ad hoc networks. In *Proc. ACM MOBIHOC 2003*, June 2003.
[3].   R. Beraldi and R. Baldoni. A Caching Scheme for Routing in Mobile Ad Hoc Networks and Its Application to ZRP. *IEEE Transactions on Computers*, Vol. 52, No. 8, Aug. 2003. pp. 1051 - 1062.
[4].   P. Jacquet, P. Muhlethaler and A. Qayyam. Optimized link-state routing protocol. *IETF MANET Internet draft*, March 2002.
[5].   C. Perkins, E. M. Royer and S. Das. Ad hoc On-Demand Distance Vector (AODV) Routing. *IETF RFC 3561*, July 2003.
[6].   D. B. Johnson, Y. Hu and D. A. Maltz. The Dynamic Source Routing Protocol (DSR) for Mobile Ad Hoc Networks for IPv4. *IETF RFC 4728*,February 2007.
[7]   H. Rangarajan and J. J. Garcia-Luna-Aceves. Using labeled paths for loopfree on-demand routing in ad hoc networks. In *Proc. ACM MOBIHOC*,May 2004.
[8].   Z. Yang, Y. Liu, and X.-Y. Li. Beyond Trilateration: On the Localizability of Wireless Ad-hoc Networks. *IEEE/ACM Transactions on Networking*,18(6) pages. 1806 - 1814, Dec. 2010.
[9].   W. Xi, Y. He, Y. Liu, J. Zhao, L. Mo, Z. Yang, J. Wang, and X.-Y. Li. Locating Sensors in the Wild: Pursuit of Ranging Quality. *ACM Sensys*,2010.
[10].   P. Bose, P. Morin, I. Stojmenovic and J. Urrutia. Routing with guaranteed delivery in ad hoc wireless networks. In *Wireless Networks*, SpringerNetherlands, Volume 7, Number 6, pages: 1572-8196 Nov. 2001.
[11].   B. Karp and H. T. Kung. Greedy perimeter stateless routing for wireless networks. In Proceedings of the ACM/IEEE International Conference on Mobile Computing and Networking *(MOBICOM)*, pages 243–254, August 2000.
[12].   X. Xiang, X. Wang, and Y. Yang, Supporting Efficient and Scalable Multicasting over Mobile Ad
          Hoc Networks. In *IEEE Transactions on Mobile Computing*, vol 10, No. 4, pp. 544-559, April 2011.
[14].   Self-Adaptive On-Demand Geographic Routing for Mobile Ad Hoc Networks Xiaojing Xiang, *Member, IEEE,* Xin Wang, *Member, IEEE,* and Zuhua Zhou, *Member, IEEE.*

Identification of Packet Dropping and Modification in Wireless Sensor Networks

[1,] B.Kishore Kumar, [2,] G.K.Venkata Narasimha Reddy

[1]*M.Tech (CSE), SJCET- Kurnool (Dt), Affiliated to JNTUA University, Andhra Pradesh, INDIA.*
[2] *Associate Professor, Department of CSE, SJCET- Kurnool (Dt), Affiliated to JNTUA University, Andhra Pradesh, INDIA.*

ABSTRACT

The Packet Droppers and Modifiers are common attacks in wireless sensor networks. It is very difficult to identify such attacks and this attack interrupts the communication in wireless multihop sensor networks. We can identify the Packet Droppers and Packet Modifiers using ranking algorithms and packet marks. The Performance is represented using detection rate and false positive probability. The Proposed scheme provides an effective mechanism for catching   compromised node.

KEYWORDS: **packet droppers and modifiers,** *intrusion detection, wireless sensor networks.*

I.   INTRODUCTION

The Simplicity in Wireless Sensor Network with resource constrained nodes makes them extremely vulnerable to variety of attacks. In a Wireless sensor networks sensor nodes monitor the environment, detect events of interest, produce data and collaborate in forwarding the data towards to a sink, which could be a gateway, base station or storage node. Securing the Wireless Sensor Networks need to make the network support all security properties: confidentiality, integrity, authenticity and availability. A sensor network is often deployed in an unattended and hostile environment to perform the monitoring and data collection tasks. When it is deployed in such an environment, it lacks physical protection and is subject to node compromise. After compromising one or multiple sensor nodes, an adversary may launch various attacks to disrupt the in-network communication. Among these attacks, two common ones are dropping packets and modifying packets, i.e., compromised nodes drop or modify the packets that they are supposed to forward.

We expect sensor networks to consist of hundreds or thousands of sensor nodes as in Fig 1. Each node represents a potential point of attack, making it impractical to monitor and protect each individual sensor from either physical or logical attack. The networks may be dispersed over a large area, further exposing them to attackers who capture and reprogram individual sensor nodes. Attackers can also obtain their own commodity sensor nodes and induce the network to accept them as legitimate nodes, or they can claim multiple identities for an altered node. Once in control of a few nodes inside the network, the adversary can then mount a variety of attacks.
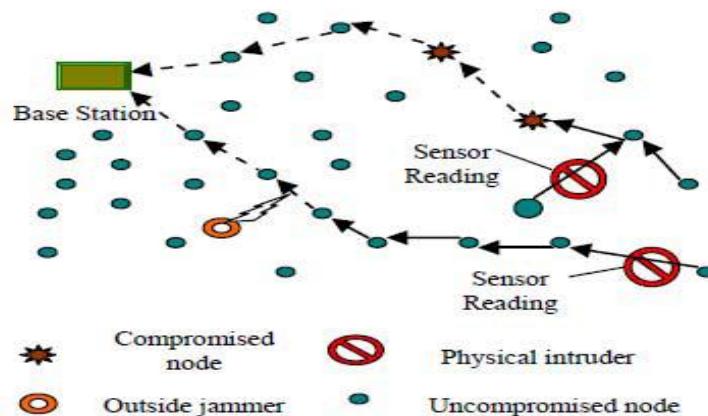
Fig 1. Sensor Network

Packet dropping is nothing but a bad node drops all or some of the packets that are supposed to be forwarded. It may also drop the data generated by itself for some malicious purpose such as blaming innocent nodes. This paper proposes a scheme to catch both packet droppers and modifiers. At first routing tree is established using DAG. Data is transmitted along the tree structure toward the sink. A packet sender or forwarder adds a small number of extra bits, which is called packet marks, is designed such that the sink can obtain the dropping ratio associated with every sensor node. Node categorization algorithm to identify nodes that are droppers/modifiers for sure or are suspicious droppers/ modifiers [1].

II.   SYSTEM MODEL

A. Network Assumptions

We assume that a typical deployment of sensor network, as where a large number of sensor nodes are deployed in a two dimensional area. Each sensor node generates sensing data periodically and all these nodes collaborate to forward packets that contain the data hop by hop towards a sink. The sink is located at some place within the network. We assume that all sensor nodes and the sink are time synchronized, which is required by many applications. The sink is aware of the network topology, which can be achieved by requiring nodes to report their neighbouring nodes soon after deployment.

B. Security Assumptions and Attack Model

We assume that the network sink is trustworthy and free of compromise, but regular sensor nodes can be compromised. Compromised nodes may or may not collude with each other. A compromised node can launch the following two attacks:

[1]   **Packet dropping:** A compromised node drops all or some of the packets that it is supposed to forward. It may also drop the data generated by itself for some malicious purpose such as accusing innocent nodes.

[2]   **Packet modification:** A compromised node modifies all or some of the packets that it is supposed to forward. It may also modify the data it generates to protect itself from being identified or to accuse other nodes.

III.   EXISTING SYSTEM

Existing counter measures aim to filter modified messages resend within a certain number of hopes. These measures can tolerate or mitigate the packet dropping and modification attacks, but the intruders are still there and can continue the network without being caught.In existing scheme, modified packets should not be filtered out en route because they should be used as evidence to infer modified packets; hence, it cannot be used together with existing packet filtering schemes.

Disadvantages of Existing system:

•   Intruders are able to collect the data while we are sending data from source to destination.
•   It is not possible to send modified packets to destination.
•   It cannot be easy to find what are the dropped and modified packets.
•   In this system, the modified packets should not be filtered out.

IV.   PROPOSED SYSTEM

Our Proposed scheme consists of system initialization phase and compromised nodes identification phases.

4.1. Initialization phase:

In the initialization phase, sensor nodes form a topology which is direct acyclic graph (DAG).A routing tree is extracted from the DAG. Data reports follow the routing tree structure. The purpose of system initialization is to set up secret pair wise keys between the sink and every regular sensor node. To establish the

Each sensor node u is preloaded the following information:
•   Ku: A secret key exclusively shared between the node and the sink.
•   Lr: The duration of a round.

- Np:The maximum number of parent nodes that each node records during the DAG establishment procedure.
- Nsth packet is numbered Ns-1,the Ns-1th packet is numbered 0,and so on and so forth.
- Ns: the maximum packet sequence number.

4.2.Intruder Identification phase:
In each round, data are transferred through the routing tree to the sink. Each packet sender/forwarder adds a small number of extra bits to the packet and also encrypts the packet. When one round finishes, based on the extra bits carried in the received packets, the sink runs a node categorization algorithm to identify nodes that must be bad nodes and suspiciously bad. The routing tree is reshaped every round, when a certain number of rounds have passed, sink collects enough information about node behaviors in different routing topologies.

4.3.Packet Sending:
when a sensor node u has a data item D to report, it composes and sends the following packet to its node.

Pu: <Pu,{Ru,u,Cp MOD Ns,D,padu,0} Ku,padu,1>

Where Pu - parent node, Ru – receiving node, U- node, Cp – counter node, D – data ,pad u,0 –padding, Ku encryption. Puddings pad u,0 and pad u,1 are added to make all packets equal in length, such that forwarding nodes cannot tell packet sources based on packet length, Meanwhile, the sink can still decrypt the packet to find out the actual content.

4.4.Packet forwarding:
When a sensor node v receives packet hv;mi, it composes and forwards the following packets to its parent node Pv:

$$<P_{v,}\{R_{v,}m\}K_v>$$

Where m is obtained by trimming the rightmost log(Np) bits off m. Meanwhile , Rv, which has logNp bits,is added to the front of m.

4.5.Packet receiving at the sink:
The sink attempts to find a child node for every parent node by decrypting which results in a string. If the attempt fails the packet is modified and it should be dropped. If it succeeds the packet is forwarded from the respective node.

4.6.Algorithm 1.Packet Receipt at the Sink
[1] Input: packet<0;m>.
[2] if Success Attempt =false then decrypt.
[3] if decryption fails then continue, else
[4] if Success Attempt=true then record sequence.
[5] u←v, Success Attempt=false;go to line4.
[6] if Success Attempt = false then
[7] drop this packet.

4.7.Algorithm 2. Tree-Based Node Categorization
1.   Input: Tree T, with each node u marked by "+" or"_," and its dropping ratio du.
2.   for each leaf node u in T find parent node until the sink node categorize the nodes.
3.   consider u as positive threshold and v as negative threshold.
4.   If v. mark ="_"then until v.mark="+" or v is Sink,Set nodes from b to e bad for sure.
5.   if v is Sink then Set u as bad for sure.
6.   If v. mark ="+" and if v is not bad for sure then set u and v as suspiciously bad else
7.   if dv – du>θ then
8.   Set v as bad for sure.
9.   if difference du-dv> θthen Set u and v as suspiciously bad;

Nu,max - most recently seen sequence number
Nu,flip   - the number of sequence number flips

nu,rcv   - number of received packets.

The dropping ratio in each round is calculated as follows:

$$d_u = \frac{N_{u,flip} * N_s + N_{u,max} + 1 - n_{u,rcv}}{N_{u,flip} * N_s + N_{u,max} + 1}$$

To identify most likely bad nodes from suspicious nodes:

$S_i = \{ <u_j, v_j> \mid <u_j, v_j> \text{ is a suspicious pair and } <u_j, v_j> = <u_j, v_j> \}$

Ranking algorithms:

1. Global ranking based approach:
        The GR method is based on the heuristic that, the more times a node is identified as suspiciously bad, the more likely it is a bad node. The node with the highest value is chosen as a most likely bad node and all the pairs that contain this node are removed.

2. Stepwise ranking based approach:
        It can be anticipated that the GR method will falsely accuse innocent nodes that frequently been parents or children of bad nodes. Once a bad node u is identified, for any other node v that has been suspected together with node u,the value of node v's accused account is reduced by the times that u and v have been suspected together.

3. Hybrid Ranking-Based(HR) Method:
        The GR Method can detect most bad nodes with some false accusations while the SR method has fever false accusations but may not detect as many bad nodes as the GR method. After a most likely bad node has been chosen, the one with the highest accused account value among the rest is chosen only if the node has not always been accused together with the bad nodes that have been identified already.

4. Packet Modifiers:

        Modified packets can be detected with the afore-described scheme. Modified packets will be detected by sink and it will be dropped and hence packet modifier can be identified as packet dropper .To enable en-route detection of modifications, the afore-described procedures for packet sending and forwarding can be slightly modified as follows. When a node *u* has a data item *D* to report , it can obtain endorsement message authentication codes (MACs) from its neighbors, which are denoted as *MAC(D)*, following existing en-route filtering schemes such as the statistical en-route filtering scheme (SEF)[3] and the interleaved hop-by-hop authentication scheme[4].

V.   SIMULATION RESULTS
5. IMPACT OF ROUND LENGTH
        Considering the delay for transmitting a packet from a source node to the sink, the round length effects number of packets received at the sink in each round, which in turn affects the detection performance. It can be seen that round length mainly affects the false positive probability. it is shown in fig. 2(a),2(b).
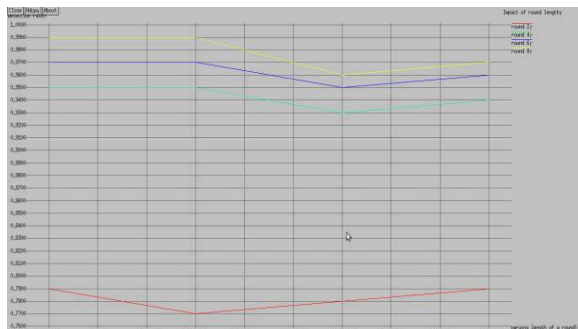
Fig 2a. Detection Rate   Fig 2b. False positive
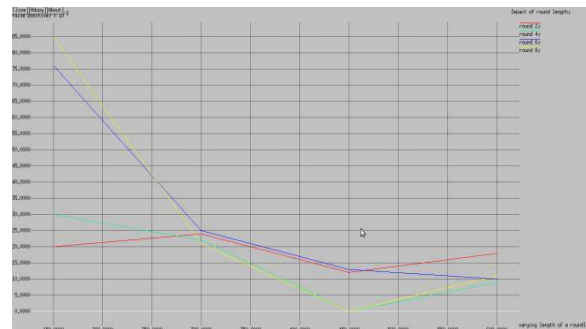
5.2.IMPACT OF DROPPING PROBABILITY

Fig.3 shows the performance sensitivity to bad node's dropping percentage (i.e., the percentage of packets that will be dropped if a bad node decides to drop packets to drop packets in a round). We vary the dropping probability between 20% and 80%.From Fig.3(a),3(b). We can see the all the three ranking algorithms have similar sensitivity to the dropping probability. In addition, with a high dropping probability, all the three algorithms achieve a higher detection rate in the early rounds, which means they can detect bad nodes quicker, and can achieve a lower false positive generally. This is because frequent misbehaviors can quickly distinguish bad nodes from innocent nodes.

Fig 3a. Dropping probability 20%-False positive   Fig 3b.Dropping probability 20%-Detection Rate
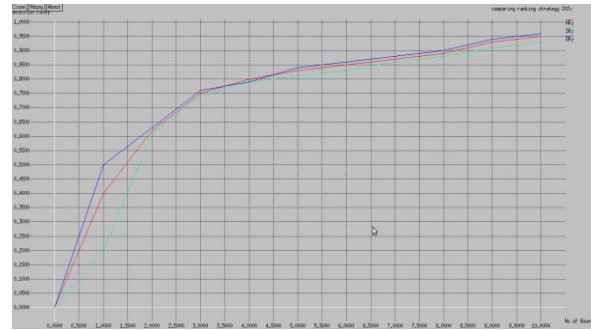
5.3.IMPACT OF THRESHOLD

Threshold for Differentiating "+" Nodes and "-" Nodes. In order to tolerate incidental packet loss, we use a threshold θ when marking each node with "+" or "-". Fig.3 shows the impact of this threshold on the detection performance. As depicted in Fig. 4(a), the larger is the threshold, the lower is the detection rate. This is because, fewer nodes will be marked as "-" as the threshold increases. Hence, a part of bad nodes may escape from being detected.

Fig.4a. Dropping Probability 80%-False positive   Fig. 4b. Dropping Probability 80%- Detection Rate
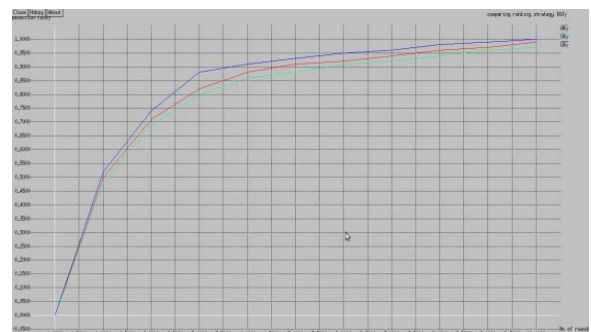
As shown in Fig. 4(b), when the threshold increases , the false positive probability increases first and then decreases after the threshold reaches a certain value (turning out).

Advantages of the system:
- A simple effective is used to catch both packet droppers and modifiers.
- While we are sending the data from source to destination, the node categorization algorithm finds the dropped and modified packets.
- Using sink node it is possible to resend the dropped and modified packets from source to destination.
- The sink can figure out the dropping ratio associated with every sensor node.

- The heuristic ranking algorithms identifies most likely bad nodes from suspiciously bad nodes.

VI. CONCLUSION

The Proposed Scheme is effective in both defecting and filtering packet droppers and modifiers. The bad nodes can be identified by the suspiciously bad nodes. The node categorization and heuristic ranking algorithms are used for this purpose. Extensive simulations have been done to prove the effectiveness of our scheme.

VII. REFERENCES

[1]. Chuang Wang, Taiming Feng, Jinsook Kim, Guiling Wang and Wensheng Zhang, Catching Packet Droppers And Modifiers In Wireless Sensor Networks, ieee transactions on parallel and distributed systems,vol.23, no. 5,may 2012..

[2]. Bhuse, A. Gupta, and L.Lilien,"DPDSN: Detection of Packet Dropping Attacks for Wireless Sensor Networks," *Proc . Fourth Trusted Internet Workshop*,2005.
Z. Yu and Y. Guan , "A Dynamic En-route Scheme for Filtering False Data in Wireless Sensor Networks,"Proc. IEEE INFOCOM,2006.

[3]. F.Ye,H.Luo,S.Lu,and L.Zhang,"Statistical En-Route Filtering of Injected False Data in Sensor Networks,"Proc.IEEE INFOCOM,2004.

[4]. S.Zhu,S. Setia,S.Jajodia,and P.Ning,"An Interleaved Hop-by-Hop Authentication Scheme for Filtering False Data in Senso Networks,"*Proc.IEEE Symp.Security and privacy* ,2004.

[5]. Issa Khalil,Saurabh Bagchi, "MISPAR: Mitigating Stealthy Packet Dropping in Locally–Monitored Multi-hop Wireless Ad Hoc Networks.

[6]. Marti, S.,Giuli, T. J., Lai ,K.,and Baker ,M., Mitigating Routing Misbehavior in Mobile Ad Hoc Networks,. Proc.6th Annual Intl.Conf.on Mobile Computing and Networking (MobiCom.00),Boston,Massachusetts,August 2000,pp 255-265.

Biomedical Image Processing With Nonlinear Filters

Himadri Nath Moulick[1] , Moumita Ghosh[2]

[1]CSE, Aryabhatta Institute of Engg & Management, Durgapur, PIN-713148, India
[2]CSE,University Institute Of Technology,(The University Of Burdwan) Pin -712104,India

ABSTRACT :

Nonlinear filtering techniques are becoming increasingly important in image processing applications, and are often better than linear filters at removing noise without distorting image features. However, design and analysis of nonlinear filters are much more difficult than for linear filters. One structure for designing nonlinear filters is mathematical morphology, which creates filters based on shape and size characteristics. Morphological filters are limited to minimum and maximum operations that introduce bias into images. This precludes the use of morphological filters in applications where accurate estimation of the true gray level is necessary. This work develops two new filtering structures based on mathematical morphology that overcome the limitations of morphological filters while retaining their emphasis on shape. The linear combinations of morphological filters eliminate the bias of the standard filters, while the value-and-criterion filters allow a variety of linear and nonlinear operations to be used in the geometric structure of morphology. One important value-and-criterion filter is the Mean of Least Variance (MLV) filter, which sharpens edges and provides noise smoothing equivalent to linear filtering. To help understand the behavior of the new filters, the deterministic and statistical properties of the filters are derived and compared to the properties of the standard morphological filters. In addition, new analysis techniques for nonlinear filters are introduced that describe the behavior of filters in the presence of rapidly fluctuating signals, impulsive noise, and corners. The corner response analysis is especially informative because it quantifies the degree to which a filter preserves corners of all angles. Examples of the new nonlinear filtering techniques are given for a variety of medical images, including thermographic, magnetic resonance, and ultrasound images. The results of the filter analyses are important in deciding which filter to use for a particular application. For thermography, accurate gray level estimation is required, so linear combinations of morphological operators are appropriate. In magnetic resonance imaging (MRI), noise reduction and contrast enhancement are desired. The MLV filter performs these tasks well on MR images. The new filters perform as well or better than previously established techniques for biomedical image enhancement in these applications.

KEY WORDS: *MLV filter , LOCO filter , pseudomedian , Erosion and the Midrange Filter Averaging Filter , Midrange Filter, (MSE) of filtered noisy signals.*

I.   INTRODUCTION

Nonlinear methods in signal and image processing have become increasingly popular over the past thirty years. There are two general families of nonlinear filters: the homomorphic and polynomial filters, and the order statisticand morphological filters [1]. Homomorphic filters were developed during the 1970's and obey a generalization of the superposition principle [2]. The polynomial filters are based on traditional nonlinear system theory and use Volterra series. Analysis and design of homomorphic and polynomial filters resemble traditional methods used for linear systems and filters in many ways. The order statistic and morphological filters, on the other hand, cannot be analyzed efficiently using generalizations of linear techniques. The median filter is an example of an order statistic filter, and is probably the oldest [3, 4] and most widely used order statistic filter. Morphological filters are based on a form of set algebra known as mathematical morphology. Most morphological filters use extreme order statistics (minimum and maximum values) within a filter window, so they are closely related to order statistic filters [5, 6].While homomorphic and polynomial filters are designed and analyzed by the techniques used to define them, order statistic filters are often chosen by more heuristic methods. As a result, the behavior of the median filter and other related filters was poorly understood for many years. In the early 1980's, important results

on the statistical behavior of the median filter were presented [7], and a new technique was developed that defined the class of signals invariant to median filtering, the root signals [8, 9]. Morphological filters are derived from a more rigorous mathematical background [10-12], which provides an excellent basis for design but few tools for analysis. Statistical and deterministic analyses for the basic morphological filters were not published until 1987 [5, 6, 13]. The understanding of the filters' behavior achieved by these analyses is not complete, however, so further study may help determine when morphological filters are best applied. This dissertation investigates the use of morphology-based nonlinear filters to enhance biomedical images. Specifically, new filters based on mathematical morphology are developed, analyzed, and applied to a variety of medical images. The behavior of the standard morphological filters is undesirable for certain applications, and the new filters are designed to overcome these weaknesses. Some new analysis techniques are introduced, including a method to quantify the response of filters to two-dimensional features. These new analysis methods and the basic statistical and deterministic analyses are used to compare the new filters with the standard filters. Finally, the new nonlinear filters are used to enhance magnetic resonance, thermographic, and ultrasound images and their performance is compared to established filtering techniques for each of the imaging modalities.

II.   ORGANIZATION

This dissertation begins with a review of mathematical morphology, including the statistical and deterministic properties of the morphological filters. These properties point out weaknesses (specifically, a bias problem) in the behavior of the standard morphological filters that motivate the development of new filters. Next, new filters that address the bias problem of the standard filters are introduced. Linear combinations of morphological operators are one of the new types of filters. This work develops the deterministic and statistical 4 proper ies of these filters and illustrates the potential advantages of these filters over the standard morphological filters. Another new type of filter introduced in this work is the value-andcriterion filter. This filter structure uses the shape-based organization of morphology, but expands the operations used for the filtering beyond just the maximum and minimum operators. Thus, any linear or nonlinear function can be used to determine the output value from values in a window, and to determine which window to use to get the output value. A promising application of this new structure is for designing filters that sharpen edges and smooth noise simultaneously. One of these new filters is the "Mean of Least Variance" filter, or MLV filter, which is a significant improvement over previously defined edgepreserving smoothing filters. The deterministic and statistical properties of the MLV filter are also investigated to contrast its behavior with other morphologybased filters. Since the usual statistical and deterministic analyses provide only an incomplete understanding of the behavior of nonlinear filters, new analysis methods are introduced here to gain further insight into the response of the filters.A technique to quantify the response of filters to periodic signals of various frequencies is outlined. This method is similar to Fourier analysis for linear filters, but is much more limited in scope because of the nonlinear nature of the filters examined. Nonetheless, this analysis gives valuable clues about the response of nonlinear filters to rapidly fluctuating signals. Another important property of many nonlinear filters is their resistance to outlying values and impulsive noise. The "breakdown point" is a measure of the robustness of filters in the presence of outliers. This method is another way to help explain differences among filters. The last analysis method developed in this dissertation furthers the understanding of the behavior of filters at two-dimensional structures. This technique, called "corner response analysis," quantifies the percentage of binary corners of various angles that is preserved by a filter. By plotting this information in polar format, the change in the response of a filter to corners of various angles is easily visualized. This method is a major improvement over previous analyses that focused on general characteristics like noise reduction or one-dimensional characteristics like edge preservation. The response of the filter to different rotations of the same feature is also explored using corner response analysis, indicating whether a filter acts similarly to different rotations of 2-D objects. The final portion of this work illustrates the use of the new nonlinear filters in biomedical image processing applications. The results for the various filters yield important information for selecting the proper filter for a given application. Among the considerations for selecting a filter are the signal and noise characteristics of the specific imaging modality and the type of information that is to be extracted from the data. The imaging modalities considered (thermography, magnetic resonance, and ultrasound) have a variety of different characteristics that call for different filters. The theoretical analyses in the earlier sections provide a solid basis for selecting appropriate filters for each modality.

III.   MATHEMATICAL MORPHOLOGY

Mathematical morphology is a set algebra used to process and analyze data based on geometric shapes. The theory of mathematical morphology was introduced by Matheron [10] in 1974 and refined by Serra [11, 12] in the 1980's. The basic morphological operations are erosion and dilation. For binary signals, erosion is a Minkowski set subtraction (an intersection of set translations), and dilation is a Minkowski set addition (a union of set translations). These operators were extended to operate on non-binary signals by Serra [11] and others [5, 15, 16].

There are two main types of morphological filters [5]: set processing and function processing filters. Set processing filters accept binary input signals and give binary output signals, while function processing filters accept binary or nonbinary functions as input and yield non-binary functions as output. The interpretation of binary signals as sets and non-binary signals as functions is straightforward; more details are given in [5].

3.1.Basic Morphological Operators
Set Processing Operations.

Let X denote an m-dimensional set and N denote a compact k-dimensional Set ($k <=m$), and let **y** denote a point in $\Re k$ and **z** a point in $\Re m$. The set X is a binary signal or image to be filtered, and the set N is called the structuring element of the morphological filter. Define the symmetric set $\tilde{N} = \{$-**z**: **z** ϵ $N\}$, which is a reflection of N about the origin. The translation of a set to a point **z** is denoted by a subscript; for example, the set N translated to the point **z** is N**z**. The set processing morphological erosion and dilation are defined by:

Erosion: $\quad X \text{ K } \tilde{N} = \{\mathbf{z}: \; N_{\mathbf{z}} \subseteq X\} = \bigcap_{\mathbf{y} \in N} X_{-\mathbf{y}}$ $\qquad$ (1)

Dilation: $\quad X \text{ Ð } \tilde{N} = \{\mathbf{z}: \; (N_{\mathbf{z}} \cap X) \neq \varnothing\} = \bigcup_{\mathbf{y} \in N} X_{-\mathbf{y}}$ $\quad$ (2)

The symbols **K** and **Ð** denote Minkowski subtraction and Minkowski addition, respectively [11, 12]. The erosion of a set X is then the set of points to which the structuring element N may be translated while remaining entirely within the original set X. The dilation of X is the set of points to which N may be translated and still intersect X with at least one point. Examples of erosion and dilation of some simple discrete sets are shown in Figure 1 below. Clearly, erosion shrinks a set while dilation expands a set.

Figure 1**.** (a) Erosion example. (b) Dilation example. (Adapted from [5].)

Squares (■) denote origin of the plane; circles (●) denote other points in the set.

In most applications of mathematical morphology, the structuring elements are symmetric about the origin, so that $N = \tilde{N}$. When this is the case, there is no distinction between erosion and Minkowski subtraction nor between dilation and Minkowski addition. Erosion and dilation also are duals of each other with respect to set complementation. If a superscript c denotes set

$$X \text{ Ð } \tilde{N} = (X^c \text{ K } \tilde{N})^c \text{ and } X \text{ K } \tilde{N} = (X^c \text{ Ð } \tilde{N})^c.$$

complementation, then

3.2.Function Processing Operations.

Since most signal and image processing applications do not deal with binary data, mathematical morphology must extend to non-binary signals (functions) to be widely useful. This extension is performed by representing a function as an ordered set of binary signals [5, 11]. The cross section of a function at a particular level is a binary set, and the set of all such cross sections forms a complete representation of the function. This process of reducing a function to a set of binary signals is called threshold decomposition. The only restriction on threshold decomposition is that the function must be upper semicontinuous, which means that each cross section of the function must be a closed set. This is not a problem in most applications because all sampled functions are upper semicontinuous [5]. Sternberg [15, 16] used another technique to extend morphology to functions, but the resulting function processing filters are identical to those derived from threshold decomposition. Threshold decomposition is illustrated in Figure 2, which shows three cross sections of a one-dimensional function $f(\mathbf{x})$. Given all the cross sections $X(f, t)$ of a function f, the original signal $f(\mathbf{x})$ may be uniquely reconstructed by simply "stacking" the cross sections. The value of the original signal at a location **x** is equal to the highest value of t for which the location **x** is included in the set $X(f, t)$. For a quantized signal, there are a finite number of levels (t) where the cross sections of the signal are taken.
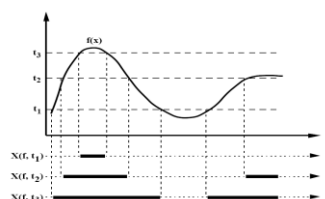
Figure 2. Example of threshold decomposition of a function into cross sections.(Adapted from [5].)

A set processing filter *f(X)* is said to be "increasing" [5] if for any two sets *A* and *B* where $A \subseteq B$, the filtered sets maintain the same set relationship; that is, $f(A) \subseteq f(B)$. This property is also called the "stacking property" [17]. A discrete, binary set processing filter possesses the stacking property if and only if its output can be expressed as a Boolean function that does not contain the complement of any of the input variables [17]. Such expressions are called positive Boolean functions. Set processing filters that are increasing (possess the stacking property) may be converted to function processing filters by performing the set processing operations on the individual cross sections of a signal; the filtered signal is found by stacking the filtered cross sections in the manner described previously. This process of converting the cross sections back into a function is a supremum operation. Examples of filters that are increasing and therefore may operate on a threshold decomposition of a signal are all order statistic filters [18, 19], including the median filter, and the morphological filters. Function processing filters that operate in this manner on the cross sections of a function are called stack filters [6, 17, 20]. Note that not all function processing filters obey the stacking property; those that do are part of a subset of function processing filters called function and set processing (FSP) filters [5, 13]. FSP filters are useful because they can accept either sets or functions as input, and give the same type of output as the input they receive. The set processing morphological filters may be converted to a function processing operation by this stacking property. The resulting grayscale morphological filters are a subset of the stack filters. All stack filters are FSP filters, and so the morphological filters that use set structuring elements are FSP filters. The resulting expressions for the FSP morphological filters are [5, 13]:

Erosion: $\quad (f \ominus \tilde{N})(\mathbf{z}) = \inf \{ f(\mathbf{y}): \mathbf{y} \in N_{\mathbf{x}} \}$ (3)

Dilation: $\quad (f \oplus \tilde{N})(\mathbf{z}) = \sup \{ f(\mathbf{y}): \mathbf{y} \in N_{\mathbf{x}} \}$ (4)

where *f(**y**)* denotes an *m*-dimensional upper semicontinuous function, *N* denotes a compact *k*-dimensional set (*k* <=*m*), and **y** and **z** denote points in $\Re k$ and $\Re m$, respectively. The infimum (inf) and supremum (sup) operations reduce to simple minimum and maximum operations, respectively, when acting on discrete signals. The structuring element of a morphological filter does not have to be a set. Like the signal the filter operates on, the structuring element may be a function. In this case, the output of the morphological filter is always a function, so such filters are not FSP, but are function processing. Instead of the structuring element set *N*, the structuring element is a *k* dimensional upper semicontinuous function *g(**z**)* that is defined over a compact region of support, *S*. As for the set structuring element case, the erosion and dilation are defined as the Minkowski subtraction and addition of the signal with the reflection of the structuring element about the origin. Let $\tilde{g}(\mathbf{z}) = g(-\mathbf{z})$ denote this reflection. The morphological erosion and dilation of a function by a function are given by:

Erosion: $\quad (f \ominus \tilde{g})(\mathbf{z}) = \inf \{ (f(\mathbf{z}+\mathbf{y}) - g(\mathbf{y})): \mathbf{y} \in S \}$ (5)

Dilation: $\quad (f \oplus \tilde{g})(\mathbf{z}) = \sup \{ (f(\mathbf{z}+\mathbf{y}) + g(\mathbf{y})): \mathbf{y} \in S \}$ (6)

Note that if the function structuring element *g(**z**)* is zero over its entire region of support *S*, then the above expressions are equivalent to the FSP expressions with structuring element *N = S*. Set structuring elements are far more widely used than function structuring elements in applications of mathematical morphology. In theory, however, function structuring elements are a significant addition to morphological filtering because they are able to process signals and image based on a specific intensity profile over a certain shape. Set structuring elements assume a flat intensity profile over their shapes.

3.3.Compound Morphological Operators

Erosion and dilation are complementary operations, one shrinking the size of objects in an image and the other expanding them. However, erosion and dilation are not inverses of each other. Some objects are completely removed by erosion, and therefore cannot be restored by dilation. Likewise, dilation often joins nearby objects which erosion cannot then separate. The compound morphological operations formed by performing the complementary operators in sequence are the morphological operations "opening" and "closing." Opening is defined as erosion followed by dilation, while closing is dilation followed by erosion. The structuring element used for the second operation is the reflection about the origin of the structuring element used for the first operation. The specific expressions for opening (denoted by a subscript) and closing (denoted by a superscript) for a set structuring element *N* acting on a function *f(**z**)* are:

Opening: $\quad f_N(\mathbf{z}) = [(f \ominus \tilde{N}) \oplus N](\mathbf{z})$ (7)

Closing: $\quad f^N(\mathbf{z}) = [(f \oplus \tilde{N}) \ominus N](\mathbf{z})$ (8)

The structuring element rotation between the erosion and dilation creates an "effective" structuring element for opening and closing that is symmetric about the origin, even if the original structuring element is asymmetric. This is illustrated in Figure 3 below. In most applications, this rotation has no practical effect since the original structuring element is usually symmetric.

Figure 3. Illustration of an "effective" structuring element of opening and closing formed by rotation between successive erosions and dilations. Squares (■) denote the origin of the plane; circles (●) denote other points in the set. The opening and closing operations are also complementary, and when applied in sequence, they form the doubly compound morphological operators open-close (OC) and close-open (CO):

Open-Closing: $$OC(f; N) = (f_N)^N \qquad (9)$$

Close-Opening: $$CO(f; N) = (f^N)_N \qquad (10)$$

IV. RESULT

Figures 4–7 below illustrate the operation of the simple and compound morphological operators on a one dimensional signal (frequency modulated in this example). For finite-length signals, there is a problem defining the output of filters near the ends of the signal; in the following examples, the output near the ends is found by "padding" the signal at either end with as many repetitions of the first or last value in the signal as necessary to define the output. Figure .4 demonstrates the effect of erosion and opening. Notice especially how opening preserves "negative" features of the signal, but cuts off positive impulses narrower than the structuring element. Figure .5 compares dilation and closing of the same signal; the effect is exactly the opposite. Closing preserves positive features and cuts off negative impulses. The effects of cascaded operators, open-closing (OC) and close-opening (CO), are shown in Figures 6 and 7 respectively. Notice that as the signal frequency increases, the close-opening tends toward flat regions with high values, whereas the open-closing yields flat regions with low values.
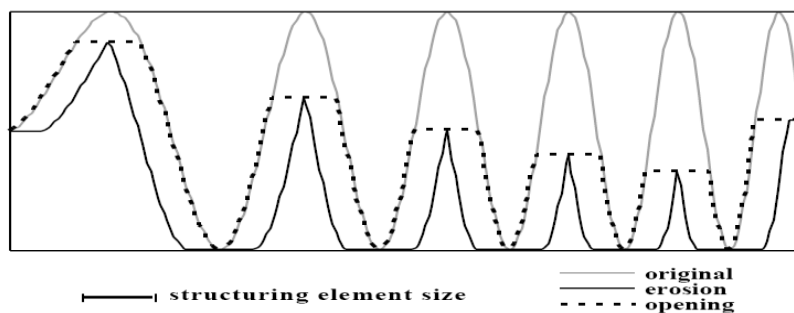
Figure 4. Effect of erosion and opening on a 1-D frequency-modulated signal.
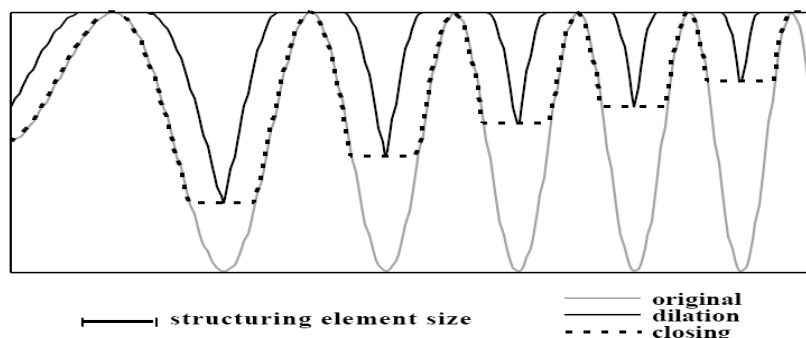
Figure 5. Effect of dilation and closing on a 1-D frequency-modulated signal.
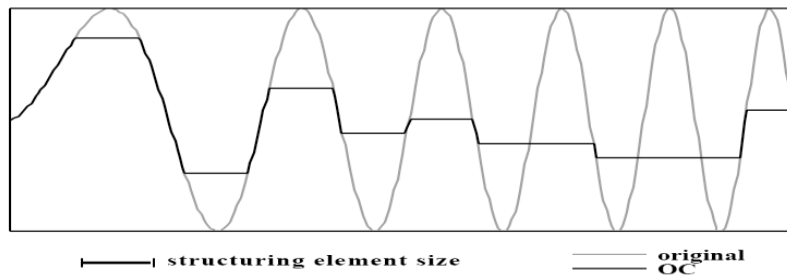
Figure 6. Effect of open-closing (OC) on a 1-D frequency-modulated signal.
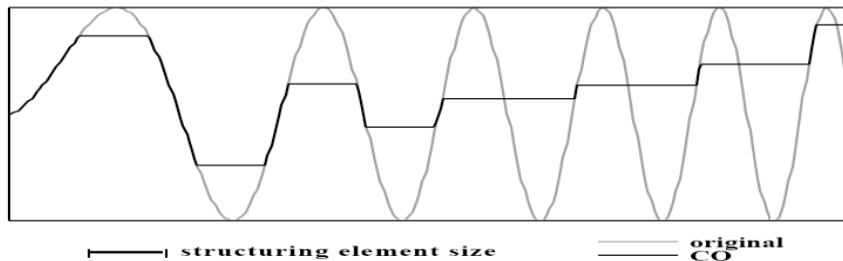
Figure 7. Effect of close-opening (CO) on a 1-D frequency-modulated signal.

Figures 8–14 below illustrate the effect of the morphological filters in two dimensions with a square structuring element. Figure 8 is the original 110 x 110 pixel binary image of several characters. In the applications given in Chapter 5, image features are considered to have high gray level values and background to have low values; this means that dilation expands image features and erosion shrinks them. However, in the binary examples given below, the characters (features) are black, and so black will be considered to take on the value 1 and white the value 0. This is inverted from the convention of 0 for black and maximum range (often 255) for white which will be used later in this work. Figures 9 and 10 show the erosion and dilation, respectively, of the original image in Figure 2.8 using a 5 x 5 square structuring element. Note that the thin parts of the characters disappear in the eroded image, while nearby portions of adjacent characters are merged in the dilated image. The opening and closing of the image with a 5 x 5 structuring element are illustrated in Figures 11 and 12, respectively. The opening, which expands the characters back from what was left after erosion, still shows many gaps where thin features were removed. The closing, however, preserves the thin features of the characters, but partially closes some of the enclosed white areas and leaves many of the characters touching. The doubly compound operators OC and CO, shown in Figures 13 and 14 respectively, are quite similar in appearance. There are only a few small differences between the OC and CO results, all of which are related to sharp corners in the original image (such as the angles in the "4"). One important trait to notice in Figures 2.9–2.14 is the effect of the structuring element shape on the filtered images. In every case, the square shape of the structuring element has a noticeable effect, preserving straight edges and 90° corners and converting other features to a "more square" appearance. The fact that the structuring element shape greatly influences the output of the filter is one of the most important features of mathematical morphology.

Figure .8. Original binary image.
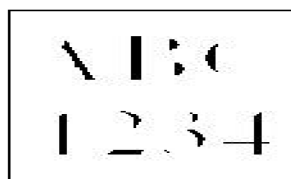
Figure 9. Image after erosion

Figure 10. Image after dilation.
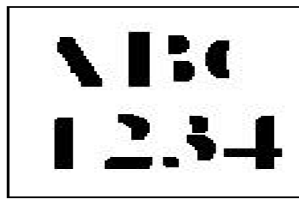
Figure 11**.** Image after opening.

Figure 12**.** Image after closing.

Figure 13**.** Image after OC.

Figure 14. Image after CO.

V.   CONCLUSION

This topic has demonstrated the utility of nonlinear filters for the enhancement of medical images, including thermography, MRI, and ultrasound. Morphology-based filters allow processing by shape and size, which can be particularly useful for feature extraction and segmentation problems in medical image analysis. For MRI, the desired processing usually includes contrast enhancement and noise reduction. These enhancements simplify the segmentation process on MR images. The morphology-based MLV filter, which is one of the new filters in the value-and-criterion structure introduced in topic 3, does an excellent job of sharpening edges and reducing noise in MR images. The MLV filter yields excellent results in only one pass, as opposed to anisotropic diffusion, which requires many iterations. The MLV filter is also computationally more efficient than similar edge-preserving smoothing filters. This work extends the theory of nonlinear image processing by introducing new filter structures and analysis methods, and demonstrates the utility of these new techniques on a variety of medical images. The filters are based on generalizations of mathematical morphology, which is itself a relatively recent development in image processing. One of these generalizations is the class of linear combinations of morphological operators. This filter class includes the previously defined midrange and pseudomedian filters and leads to the definition of the LOCO filter. The deterministic properties of the linear combinations are similar to those of the constituent morphological filters; however, the linear combinations are statistically unbiased, unlike the conventional morphological operators. This is important in applications like thermography, where the LOCO filter allows shape-based filtering as in standard morphology without introducing a statistical bias. The other new class of filters introduced here is the value-and-criterion filter structure. This structure is based on morphological opening and closing, but allows the use of both linear and nonlinear operators within the window structure. One useful filter designed with this structure is the Mean of Least Variance (MLV) filter, which takes the mean of the window with the smallest variance within a set of windows defined by the morphological structure. This filter reduces noise and enhances edges in images. Its statistical and deterministic properties resemble those of the averaging filter more than the morphological filters, with several important differences. The MLV filter is valuable in applications like MRI, where noise smoothing and contrast enhancement are the primary goals.

REFERENCE

[1]   I. Pitas and A. N. Venetsanopoulos, Nonlinear Digital Filters: Principles and Applications. Boston: Kluwer Academic, 1990.
[2]   A. V. Oppenheim and R. W. Schafer, Discrete-Time Signal Processing. Englewood Cliffs, New Jersey: Prentice Hall, 1989.
[3]   J. W. Tukey, Exploratory Data Analysis. Reading, Massachusetts: Addison-Wesley, 1971.
[4]   R. P. Borda and J. D. Frost, Jr., "Error reduction in small sample averaging through the use of the median rather than the mean," Electroencephalography and Clinical Neurophysiology, vol. 25, pp. 391- 392, 1968.
[5]   P. Maragos and R. W. Schafer, "Morphological filters—Part I: Their settheoretic analysis and relations to linear shift-invariant filters," IEEE Trans. Acoust., Speech, Signal Process., vol. 35, no. 8, pp. 1153-1169, 1987.
[6]   P. Maragos and R. W. Schafer, "Morphological filters—Part II: Their relations to median, order-statistic, and stack filters," IEEE Trans. Acoust., Speech, Signal Process., vol. 35, no. 8, pp. 1170-1184, 1987.
[7]   B. I. Justusson, "Median filtering: Statistical properties," in Two- Dimensional Digital Signal Processing II: Transforms and Median Filtering, T. S. Huang, Editor. Berlin: Springer-Verlag, 1981. pp. 161- 196.
[8]   S. G. Tyan, "Median filtering: Deterministic properties," in Two- Dimensional Digital Signal Processing II: Transforms and Median Filtering, T. S. Huang, Editor. Berlin: Springer-Verlag, 1981. pp. 197- 217.
[9]   N. C. Gallagher, Jr. and G. L. Wise, "A theoretical analysis of the properties of the median filter," IEEE Trans. Acoust., Speech, Signal Process., vol. 29, no. 6, pp. 1136-1141, 1981.
[10]   G. Matheron, Random Sets and Integral Geometry. New York: Wiley, 1974.
[11]   J. Serra, Image Analysis and Mathematical Morphology, Vol. 1. London: Academic, 1982.
[12]   J. Serra, Editor. Image Analysis and Mathematical Morphology, Vol. 2: Theoretical Advances. London: Academic, 1988.
[13]   R. L. Stevenson and G. R. Arce, "Morphological filters: Statistics and further syntactic properties," IEEE Trans. Circuits *Syst.*, vol. 34, no. 11, pp. 1292-1305, 1987.

[14]   G. Gerig, O. Kübler, R. Kikinis, and F. A. Jolesz, "Nonlinear anisotropic filtering of MRI data," *IEEE Trans. Med. Imag.*, vol. 11, no. 2, pp. 221- 232, 1992.
[15]   S. R. Sternberg, "Biomedical image processing," IEEE Computer, vol. 16, no. 1, pp. 22-34, 1983.
[16]   S. R. Sternberg, "Grayscale morphology," Comp. Vision, Graphics, Image Process., vol. 35, pp. 333-355, 1986.
[17]   P. D. Wendt, E. J. Coyle, and N. C. Gallagher, "Stack filters," IEEE Trans. Acoust., Speech, Signal Process., vol. 34, no. 4, pp. 898-911, 1986.
[18]   J. P. Fitch, E. J. Coyle, and N. C. Gallagher, "Median filtering by threshold decomposition," IEEE Trans. Acoust., Speech, Signal Process., vol. 32, pp. 1183-1188, 1984.
[19]   J. P. Fitch, E. J. Coyle, and N. C. Gallagher, "Threshold decomposition of multidimensional ranked order operations," IEEE Trans. Circuits Syst., vol. 32, pp. 445-450, 1985.
[20]   E. J. Coyle and J.-H. Lin, "Stack filters and the mean absolute error criterion," IEEE Trans. Acoust., Speech, Signal Process., vol. 36, no. 8, pp. 1244-1254, 1988.
[21]   H. A. David, Order Statistics, 1st ed. New York: Wiley, 1970.
[22]   R. V. Hogg and A. T. Craig, Introduction to Mathematical Statistics. 4th ed. New York: Macmillan, 1978.
[23]   L. Koskinen, J. Astola, and Y. Neuvo, "Morphological filtering of noisy images," in Visual Communications and Image Processing *'90*, M. Kunt, Editor, *Proc. SPIE*, vol. 1360, pp. 155-165, 1990.
[24]   J. Astola, L. Koskinen, and Y. Neuvo, "Statistical properties of discrete morphological filters," in Mathematical Morphology in Image Processing, E. R. Dougherty, Editor. New York: Marcel Dekker, 1993. pp. 93-120.
[25]   H. A. David, Order Statistics, 2nd ed. New York: Wiley, 1980.
[26]   A. C. Bovik, T. S. Huang, and D. C. Munson, Jr., "A generalization of median filtering using linear combinations of order statistics," IEEE Trans. Acoust., Speech, Signal Process., vol. 31, no. 6, pp. 1342-1350,1983.
[27]   A. Restrepo and A. C. Bovik, "Adaptive trimmed mean filters for image restoration," IEEE Trans. Acoust., Speech, Signal Process., vol. 36, no. 8, pp. 1326-1337, 1988.
[28]   W. K. Pratt, T. J. Cooper, and I. Kabir, "Pseudomedian filter," in Architectures and Algorithms for Digital Image Processing II, F. J. Corbett, Editor, *Proc. SPIE*, vol. 534, pp. 34-43, 1985.
[29]   W. K. Pratt, Digital Image Processing, 2nd ed. New York: Wiley, 1991.
[30]   M. A. Schulze and J. A. Pearce, "Continuous time analysis of the response of the pseudomedian and related filters to periodic signals," in Nonlinear Image Processing *III*, E. R. Dougherty, J. Astola, and C. G. Boncelet, Jr., Editors, *Proc. SPIE*, vol. 1658, pp. 177-188, 1992.
[31]   M. A. Schulze, Mathematical Properties of the Pseudomedian Filter. M.S. Thesis, University of Texas at Austin, 1990.
[32]   M. A. Schulze and J. A. Pearce, "Some properties of the two-dimensional pseudomedian filter," in Nonlinear Image Processing II, E. R. Dougherty, G. R. Arce, and C. G. Boncelet, Jr., Editors, *Proc. SPIE*, vol. 1451, pp. 48-57, 1991.
[33]   J. Song and E. J. Delp, "A study of the generalized morphological filter," *Circuits, Syst., Signal Process.*, vol. 11, no. 1, pp. 227-252, 1992.
[34]   M. A. Schulze and J. A. Pearce, "Linear combinations of morphological operators: The midrange, pseudomedian, and LOCO filters," in *Proc.* 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. V, pp. 57-60, 1993.
[35]   E. J. Gumbel, Statisti*c*s of Extremes. New York: Columbia University Press, 1958.
[36]   J. Y. Hsiao and A. A. Sawchuk, "Supervised textured image segmentation using feature smoothing and probabilistic relaxation techniques," IEEE Trans. Patt. Anal. Mach. Intell., vol. 11, no. 12, pp. 1279-1292, 1989.
[37]   M. Kuwahara, K. Hachimura, S. Eiho, and M. Kinoshita, "Processing of RI-angiocardiographic images," in Digital Processing of Biomedical Images, K. Preston, Jr. and M. Onoe, Editors. New York: Plenum, 1976.pp. 187-202.
[38]   F. Tomita and S. Tsuji, "Extraction of multiple regions by smoothing in selected neighborhoods," IEEE Trans. Syst., Man, Cyber., vol. 7, pp. 107- 109, 1977.
[39]   M. Nagao and T. Matsuyama, "Edge preserving smoothing," Comp. Graphics Image Process., vol. 9, pp. 394-407, 1979.
[40]   J.-S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Trans.* Patt. Anal. Mach. Intell., vol. 2, pp. 165-168, 1980.
[41]   J.-S. Lee, "Refined filtering of image noise using local statistics," Comp. Graphics Image Process., vol. 15, pp. 380-389, 1981.
[42]   J. Neejärvi, P. Heinonen, and Y. Neuvo, "Sine wave responses of median type filters," in Proc. 1988 IEEE International Symposium on Circuits and Systems, pp. 1503-1506, 1988.
[43]   J. Neejärvi and Y. Neuvo, "Sinusoidal and pulse responses of morphological filters," in Proc. 1990 IEEE International Symposium on Circuits and Systems, vol. 3, pp. 2136-2139, 1990.
[44]   J. P. Fitch, E. J. Coyle, and N. C. Gallagher, "The analog median filter," IEEE Trans. Circuits Syst., vol. 33, no. 1, pp. 94-102, 1986.
[45]   F. R. Hampel, "The breakdown points of the mean combined with some rejection rules," Technometric*s*, vol. 27, no. 2, pp. 95-107, 1985.
[46]   D. F. Andrews, P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, and J. W. Tukey, Robust Estimates of Location: Survey and Advances. Princeton: Princeton University Press, 1972.
[47]   H. G. Longbotham and N. Barsalou, "The LMS, an adaptive optimal order statistic filter," in Nonlinear Image Processing, E. J. Delp, Editor, *Proc. SPIE*, vol. 1247, pp. 78-88, 1990.
[48]   J. B. Bednar and T. L. Watt, "Alpha-trimmed means and their relationship to median filters," IEEE Trans. Acoust., Speech, Signal Process., vol. 32, no. 1, pp. 145-153, 1984.
[49]   C. G. Boncelet, Jr., R. Hardie, R. Hakami, and G. R. Arce, "LUM filters for smoothing and sharpening," in Nonlinear Image Processing II, E. R. Dougherty, G. R. Arce, and C. G. Boncelet, Jr., Editors, *Proc. SPIE*, vol.1451, pp. 70-74, 1991.
[50]   A. Rosenfeld and A. C. Kak, Digital Picture Processing, 2nd ed. Vol. 1. Orlando: Academic, 1982.
[51]   A. C. Bovik, T. S. Huang, and D. C. Munson, "The effect of median filtering on edge estimation and detection," IEEE Trans. Patt. Anal. Mach. Intell., vol. 9, pp. 181-194, 1987.
[52]   H. Longbotham and D. Eberly, "The WMMR filters: A class of robust edge enhancers," IEEE Trans. Signal Process., vol. 41, no. 4, pp. 1680- 1685, 1993.
[53]   Z. M. Ryu, The Effect of Nonlinear Filtering on the Resolution of Calibrated Thermal Images. Ph.D. Dissertation, University of Texas at Austin, 1986.
[54]   C.-H. H. Chu and E. J. Delp, "Impulsive noise suppression and background normalization of electrocardiogram signals using morphological operators," IEEE Trans. Biomed. *Eng.*, vol. 36, no. 2, pp. 262-273, 1989.
[55]   P. E. Trahanias, "An approach to *QRS* complex detection using mathematical morphology," IEEE Trans. Biomed. Eng., vol. 40, no. 2, pp. 201-205, 1993.

[56]   J. W. Klingler, Jr., C. L. Vaughan, T. D. Fraker, Jr., and L. T. Andrews, "Segmentation of echocardiographic images using mathematical morphology," IEEE Trans. Biomed. Eng., vol. 35, no. 11, pp. 925-934,1988.

[57]   J. G. Thomas, I. Peters R. A., and P. Jeanty, "Automatic segmentation of ultrasound images using morphological operators," IEEE Trans. Med. Imag., vol. 10, no. 2, pp. 180-186, 1991.

[58]   C. Lamberti and F. Sgallari, "Anisotropic diffusion and morphological approaches for echocardiography image processing," in *Proc.* Fifth European Signal Processing Conference, L. Torres, E. Masgrau, and M. A. Lagunas, Editors. Amsterdam: Elsevier Science, 1990. pp. 939-942.

[59]   B. D. Thackray and A. C. Nelson, "Semi-automatic segmentation of vascular network images using a rotating structuring element (ROSE) with mathematical morphology and dual feature thresholding," IEEE Trans. Med. Imag., vol. 12, no. 3, pp. 385-392, 1993.

[60]   J. Samarabandu, R. Acharya, E. Hausmann, and K. Allen, "Analysis of bone X-rays using morphological fractals," IEEE Trans. Med. Imag., vol. 12, no. 3, pp. 466-470, 1993.

Design of 132/33KV Substation

[1,]Sudipta Sen , [2,]Arindam Chatterjee , [3,]Debanjan Sarkar

[1](Electrical Engineering, Techno India/West Bengal University of Technology, India)
[2](School of Mechanical and Building Sciences, Vellore institute of technology (VIT), India)
(Electrical Engineering, Techno India/ West Bengal University of Technology, India)

ABSTRACT

The project work assigned to us was to design a 132/33 KV EHV sub-station. We considered incoming power at 132 KV and the power was transferred to main bus through isolator-circuit breaker-isolator combination. The power from main bus was fed into a 20MVA transformer which stepped the voltage down to 33KV. The power is then fed into a 33KV bus from which different loads were tapped. In the process, the surge impedance loading of 132 KV and 33 KV lines were calculated and they were used to estimate the maximum power that can be transferred by one transmission line. The design of the entire substation was made keeping in mind the most basic requirements of a proper substation including the civil and domestic requirements. The design is then submitted to our mentor for verification

KEYWORDS: *1) Bus bar 2) Control Cable 3) Earthing 4) Insulation-Coordination 5) Insulator 6) Isolator 7) Lightning Arrester 8) Power Transformer 9) Sub-Station 10) Switchgear*

I.   INTRODUCTION

The work designated to the students was to design a 132KV/33KV EHV sub-station. The work was carried out under Prof. S. Pal, H.O.D.- EE dept., Techno India. Any sub-station which handles power at over 33KV is termed as extra High Voltage sub-station by the rules implemented by Indian government. The design process of an EHV sub-station begins with very elemental work of selection of site and estimation of requirements which includes capital and material. It is also needed to keep in mind, the civil aspects of a sub-station design.In India about 75% of electric power used is generated in thermal and nuclear plants, 23% from mostly hydro station and 2% comes from renewable and other resources. The distribution system supplies power to the end consumer, while the transmission system connects between the generating stations and distribution system through transmission line. The entire network forms a power grid and each power grid across the country is interconnected which facilitates uninterrupted supply.

While designing a power grid the following aspects must be taken into consideration:

- ✓ Low capital cost.
- ✓ Reliability of the supply power.
- ✓ Low operating cost
- ✓ High efficiency
- ✓ Low cost of energy generation.
- ✓ Simplicity of design.
- ✓ Reserve capacity to meet future requirements

Starting from the generating stations to the end users, voltage is needed to be stepped up and down several times in various substations. This ensures efficient transmission of power, minimizing the power losses. Our project is to design a 132KV/33KV EHV sub-station where the incoming power is received at 132 KV from a generating station. The power factor is corrected here and the voltage is stepped down to 33KV and power is then transferred to distribution system of the grid to meet the requirements of the end consumers at their suitable voltage.

II.   A DESIGN LAYOUT OF 132/33 KV, 200 MW SUB-STATION

The sub-station is connected with three substations or load viz. A (3.2 mw), B (3.2MW) and C (3.2MW) at 33KV and D (36MW) at 132 KV. The generated 16.2 KV is stepped up to 132 KV and is supplied

to the 132KV sub-station through two double circuit transmission lines. After analyzing the requirements of the loads & SIL of transmission lines the whole arrangements are done in the following way:

2.1 Assumptions
* The value of surge impedance of transmission lines under consideration = 325 Ω
* Total load requirement = 3.2 MW + 3.2 MW + 3.2 MW + 36 MW
* The distance between the substation & the neighboring generating station is 50km.

The SIL of 132 KV line = $(132KV)^2/325$ = 53.61 = 54 MW (approx)

The SIL of 33 KV line = $(33KV)^2/325$ = 3.35 = 3.5 MW (approx)

Observing the total load demand, the input to the substation must be greater than the requirement. So one double circuit 132KV transmission lines (54 X 2 = 108 MW) only can satisfy this. The second double circuit tower is constructed keeping in mind the future load demand increase. The lines first supply the power to the 132KV bus A of the sub-station. As the distance between the substation and the generating station is only 50km, the SIL can increase to 1.2 times of the theoretical value. Hence the input of the substation can be as high as (108 X 1.2) MW i.e. almost 130 MW.
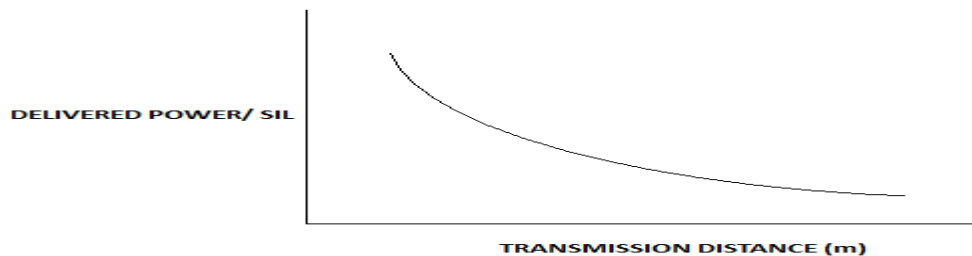
THE CHARACTERISTICS CURVE OF DELIVERED POWER/SIL Vs. TRANSMISSION DISTANCE

(The curve is closely applicable in determining transmission line loading based on transient stability & also steady state stability for operating voltages between 66 & 500KV) For load A, B and C it is suitable to step down the incoming 132KV voltage to 33KV. Hence power transformers of rating (132/33KV, 20 MVA are used). Another transfer of same rating is installed to meet future increase in demand. On the other hand, a double circuit line from 132 KV bus A is useful to serve the load D.33 KV is supplied to load A, B and C through one double circuit transmission lines (SIL capacity 3.5 X 2 = 7 MW) and to load D through one double circuit transmission lines(SIL capacity 54 X 2 = 108 MW) where actually one circuit will be left for emergency or maintenance reason.The stepped down 33KV is further stepped down to 11KV and then finally to 440V to meet the demand of local station loads.A transfer bus is installed in the system for providing provision for maintenance of the main bus.
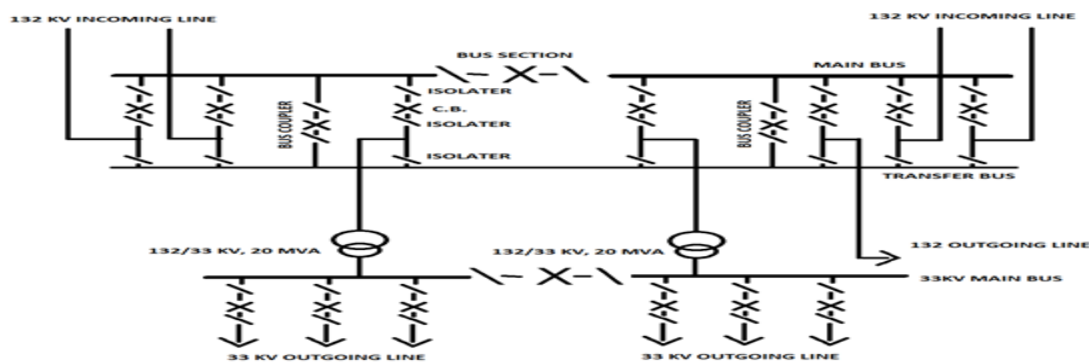
SUB STATION LAYOUT

III. DESIGN OF EARTH-MAT

3.1.Calculation

Fault current   = 40KA

Fault duration   = 1.0 sec

Soil resistivity   = 10 ohm

Depth of burial   = 0.6 M

Earth electrode   = 40 mm dia. G.I. pipe, 3 M long

Earth mat conductor = M.S. Round

Riser   = G.I. strip

Minimum cross-section of grounding conductor having required thermal stability can be calculated by using the formula,

Amin = required conductor section

If =fault current in Amps

t = time in sec for operation of protection relay

c = constant which is equal to 70 for M.S.round

Hence Amin = (4000x√1)/70=571 mm^2

Next standard size M.S. round =32 mm (diameter)

Considering soil resistivity for conductor sizing as 10 ohm-M, correction factor is taken as 1.3

   Hence cross-section area of each conductor with correction = 1.3x571 mm^2 = 742 mm^2

 Or, (∏/4)*(dia. Of conductor) = 742 mm^2

 Or, diameter of the conductor = 30.74 mm^2

Nearest standard size is 32 mm diameter

For riser connection above ground, no tolerance is required.

Hence selected size of M.S flat = 75x8 mm

Calculation of Tolerable Touch & Step Potential

The reduction factor Cs can be approximately by the equation,

Cs = 1-[0.9(1-(P/Ps))/(2hs+0.09)]

Where, P = soil resistivity   = 10 diameter

       Ps = surface layer resistivity = 2500 ohm-m

       hs = surface layer thickness   = 0.1 meter

Hence, Cs = 1-[{0.9x(1-(10/2500))}/(2x0.1+0.09)]

       = 1-(0.08964/0.2)

=0.691

Following operation can be used to compute the tolerable touch and step voltages respectively:

Etouch = (1000+1.5xCs.Ps)x0.116/√ts

Where, ts =duration of shock for determining allowable body current

= 1 sec.

Hence, Etouch = (1000+1.5x0.691x2500)x0.116/√1

= 416.59 volts

Estep = (1000+6.0xCsxPs)x0.116/√ts

= (1000+6.0x0.691x2500)x0.116/√1

=1318.34 volts

3.2. Determination of grid resistance:

The equivalent length of earth-mat area (L) = 300M

The equivalent width of earth-mat area (W) = 250M

No. of conductors along length (NL)   = 16

No. of conductors along width (NW)   = 20

Minimum no of electrodes = fault current/500= 80

Keeping a margin of 50% extra, no. of electrodes (N) = 1.5x80= 120

Length of individual electrode (Lr) = 3

Hence, LT=Lc+LR=(LxNL+WxNW)+(NrxLr)

Or, LT = (3000x16+250x20) + (120x3)

= 10160 m

Total area of earth-mat (A) = 75,000 m^2

3.3. Safety Check

For the safe earthing design, attainable step and touch voltage should be less tolerable values respectively.

Volt	Attainable	Tolerable
Touch Voltage	12.5	416.59
Step Voltage	36.72	1318.34

The attainable touch as well as step voltage is well below tolerable limit. Hence the design is safe

3.4.Earthing
Earthing means that, making a connection to the general mass of the earth. The use earthing is so widespread in an electrical system that at particular every point in the system, from the generators to the consumer equipment, earth connections are made.
The subject of earthing may be divided into
1.1   Neutral Earthing
1.2   General Earthing

IV.   INSULATION COORDINATION
Insulation co-ordination is the process of determining the proper insulation levels of various components in a power system as well as their arrangements. It is the selection of an insulation structure that will withstand voltage stresses to which the system or equipment will be subjected to, together with the proper surge arrester. The process is determined from the known characteristics of voltage surges and the characteristics of surge arrester. Its final objective is to ensure safe, optimized distribution of electric power. By optimized is meant finding the best possible economic balance between the various parameters depending on this co-ordination: n cost of insulation, n cost of protective devices, n cost of failures in view of their probality.

V.   DESIGN OF BUS BARS
Bus bars are Cu/Al rods of thin walled tubes and operate at constant voltage. The bus-bars are designed to carry normal current continuously. The cross section of conductors is designed on the basis of rated normal current and the following factors: System voltage, position of sub-station. Flexibility, reliability of supply and cost. Our design must ensure easy and uninterrupted maintenance, avoiding any danger to the operating of operating personnel. It must be simple in design and must possess provision for future extension. Any fluctuation of load must not hinder its mechanical characters. The sub-station bus bars are broadly classified in the following three categories:

1.3   Outdoor rigid tubular bus-bars.
1.4   Outdoor flexible ACSR or Al alloy bus-bars.
1.5   Indoor bus bars.

In our substation, we have chosen ONE MAIN BUS AND ONE TRANSFER BUS system. The buses are coupled using a bus-coupler which facilitates load transfer while maintenance and fault conditions.

Load catered = 200 MW

Voltage = 132 KV

Rated current is taken to be I ampere, we get

$P = \sqrt{3}\ VI \cos \varphi$

We take power factor as 0.9

= 971.97 ampere

Going by the rated current that is required to be catered and keeping in mind the future provision of load we chose twin moose conductor for the purpose of main bus and normal single moose and normal moose for transfer bus.

VI.   INSULATORS
The insulators serve two purposes. They support the conductors (or bus-bars) and confine the current to the conductors. The most commonly used material for the manufacture of insulator is porcelain. There are several type of insulators, and their use in the sub-station will depend upon the service requirement. The main four types of insulators are as follows:
8.1   Pin Type Insulators
8.2   Suspension Type Insulators

8.3   Strain Insulators
8.4   Shackle Insulators

VII.   CIRCUIT BREAKER

Circuit breakers are a piece of electrical device that
9.1 Make or break a circuit either manually or by remote control under normal conditions.
9.2 Break a circuit automatically under fault conditions.
9.3 Make a circuit either manually or by remote control under fault conditions.

Classification of Circuit Breakers:
        The most common method of classification of circuit breakers is on the basis of medium used for arc extinction. Accordingly they are classified as:

1.5.1   Oil circuit-breaker.
1.5.2   Air-blast circuit breaker.
1.5.3   Sulphur hexafluoride circuit breakers.
1.5.4   Vacuum circuit breakers.

Relays
        A protective relay is a device that detects the fault and initiates the operation of the circuit breaker to isolate the defective element from the rest of the system. The relay constantly measures the electrical quantities which are different under normal and fault condition. Having detected the fault the relay operates to close the trip circuit of the breaker. The trip circuit is operated by a direct voltage. A relay must be highly selective to the normal and fault conditions to avoid unwanted tripping. It must operate with suitable speed so that fault is eliminated before it can cause any damage. A relay must also be sensitive to work with low values of currents.

Classification of Relay
a. Electromagnetic attraction type- which operates on the principle where the relay armature is
     attracted by an electromagnet.
b. Electromagnetic induction type- which operates due to mutual interaction of two different fluxes which are
     differing at a certain phase angles, having same or different amplitude and nearly equal frequencies. The net
     torque that operates to rotate the induction disc is proportional to the product of the amplitudes and sine of
     the phase diff

Functional Relay Types
[1]   Induction type over-current relay
[2]   Induction type reverse power relay
[3]   Distance or Impedance relay
[4]   Differential relay
[5]   Translay scheme

BUCHHOLZ RELAY
        It is a gas actuated relay installed in a oil immersed transformers for protection against all kinds of faults. This relay is used to give an alarm in case of incipient (slow developing) faults in transformer and to disconnect the transformer from the supply in the event of severe internal faults. It is usually connected in the pipe connecting the conservator to the

Main Tank.

CURRENT TRANSFORMER

CT has a primary winding one or more turns of thick wire connected in series with the line carrying the current to be measure. The secondary consist of a large no of turns of fine wire and feeds a standard 5 amp. ammeter. It is used for the measuring and protection purpose. The secondary of current transformer should never be left open under any circumstances.

POTENTIAL TRANSFORMER

These transformers are extremely accurate ratio step down transformer s and are used in conjunction with standard low range voltmeter (100-120V) whose deflection when divided by transformation ratio, gives the true voltage on primary side. In general they are shell type. Their rating is extremely small for safety operation secondary is completely insulated from high voltage primary. Its primary current is determined by the load on secondary.

Lightning Arrester

An electric discharge between clouds and earth, between clouds or between the charge centers of the same cloud is known as lightning. It is a huge spark and takes place when the clouds are charged to such high potential with respect to earth or a neighboring cloud that the dielectric strength of neighboring medium is destroyed. A lightning may strike the power system (e.g. overhead lines, towers or sub-stations) directly and the current path may be over the insulators down to pole to the ground or it may strike indirectly, resulting from electrostatically induced charges on the conductors due to the presence of charged clouds.

Types of Lightning Arresters

The lightning arrester mainly differs in their constructional features. However they work with the same operating principle, i.e. providing low resistance path for the surges. They are mainly classified as:

[1]   Rod gap arrester
[2]   Horn gap arrester
[3]   Multigap arrester
[4]   Expulsion type lightning arrester
[5]   Valve type lightning arrester

VIII. SWITCHGEAR

The term switchgear, used in association with the electric power system, or grid, refers to the combination of electrical disconnects, fuses and/or circuit breakers used to isolate electrical equipment. Switchgear is used both to de-energize equipment to allow work to be done and to clear faults downstream. Switchgear is already a plural, much like the software term code/codes, and is never used as switchgears. The very earliest central power stations used simple open knife switches, mounted on insulating panels of marble or asbestos. Power levels and voltages rapidly escalated, making open manually-operated switches too dangerous to use for anything other than isolation of a de-energized circuit. Oil-filled equipment allowed arc energy to be contained and safely controlled. By the early 20th century, a switchgear line-up would be a metal-enclosed structure with electrically-operated switching elements, using oil circuit breakers. Today, oil-filled equipment has largely been replaced by air-blast, vacuum, or SF6 equipment, allowing large currents and power levels to be safely controlled by automatic equipment incorporating digital controls, protection, metering and communications.

Switchgear Panel

Power Transformer

This is the most important component of a sub-station. The main work of a sub-station is to distribute power at a low voltage, by stepping down the voltage that it receives in its incoming lines. Power is generally transmitted over long distances at very high voltages, generally in the range of 400 KV, 200 KV or 132 KV to the sub-stations. However consumer requires power at rather low voltages, 11 KV for industries and 440 V or 220 V for domestic consumers. The sub-stations use step-down transformers to attain this voltage and then distribute this power. All the other equipment in a sub-station works only to facilitate the smooth working of the power transformer.

Control Cable

Control cables are used in substations for connecting control systems, measurements, signaling devices, protection circuits etc. rated below 1000volts. They have a copper conductor. They may have another rubber insulation or PVC insulation. Control cables have several cores, each having independent insulation.To avoid interference due to stray magnetic fields, the control cables should be properly laid and their sheath should be properly earthed.

Design of Control and Relay Panel Complete with Protection for 132/33 KV Sub station

The scope of this section covers design, engineering, manufacture, installation, testing and commissioning of control and relay panels (Complete with protective relays, measuring and indicating equipments along with visual and audible alarm, interlocking schemes) inclusive of internal wiring and external connection to various switchyard equipments.

In a 132/33 KV substation the control panels are corridor type (also called duplex type). In this type the front and rear walls are erected independent with a common cover. The sides are open except the end panels, which are provided with doors and door switch for internal illumination. In between front and rear there is adequate space to move for inspection and wiring. In this type, the protective relays are mounted on rear board and the control and indication equipments on the front panels.The standard size of individual panel is Depth – 1983mm, width- limited to 1000mm, height – 2312 mm. the corridor is 762mm wide and access doors on end panels are 1900mm high. Panels are dust, moisture and vermin proof. These are free standing, floor mounting type but grounded with foundation bolts. Cable entries to panel are from bottom. The bottom plates of the panel are fitted with removable gland plates and fixed with cable glands. The cable glands are screwed type made of brass and suitable for PVC armored cable.

Mimic Diagram

Colored mimic diagram and symbols showing exact representation of the system are provided in the front panel. Mimic diagram are made of anodized aluminum or plastic, screwed to panel. The mimic buses are generally 2mm thick; width of mimic bus is 10mm for bus bar and 7mm for other connections. Indicating lamp, one for each phase for each bus is provided on mimic of bus coupler panel to indicate bus charged condition.

Color scheme for mimic diagram

VOLTAGE CLASS	COLOUR	SHADE INDEX OF IS
132KV	LIGHT ORANGE	557
33KV	SIGNAL RED	537

Semaphore indicators for each earth switches, control switch width on, off indicating lamps for isolators, and discrepancy type control switch with built-in indicating lamp, flush type for circuit breakers are mounted along mimic diagram at appropriate location in panel.

DC Circuit

There shall be only one DC incoming (220V) for the C&R board through a 32AMP switch-fuse unit. One HRC fuse-unit both at positive and negative side shall be provided for the DC incomer at the bus coupler panel. The said DC incoming bus shall run continuously in the total C&R board. DC annunciation bus shall also be teed off from the incomer DC bus through 6A HRC fuse at positive and a link in the negative side with necessary DC supervision relay.
DC supply to each individual panel thus teed off and distributed within the panel as below

[1]   C.B. remote and local closing through HRC fuse and link.
[2]   C.B. remote and protection trip to trip coil 1 with trip circuit supervision relay through a separate HRC fuse and link.
[3]   C.B. remote and protection trip to trip coil 2 with trip circuit supervision relay through a separate HRC fuse and link.
[4]   Protective relay and PT selection circuit with DC supervision relay.
[5]   Indication circuit through 6A HRC fuse and link.
[6]   Isolator control circuit through 10A HRC fuse and link.

Bus bar protection and LBB protection DC shall be teed off from the 132KV C&R board.

AC Circuit

A 132KV single phase AC supply to the entire C&R board will be fed from AC distribution board through a 32A switch-fuse unit. The supply shall be provided in bus coupler/ bus transfer panel.AC circuit for incoming DC and annunciation DC fail alarm scheme is provided in bus coupler panel. The above bus is teed off to each panel through separate switch-fuse unit. One supervision relay for incoming AC fail with test push button and reverse flag indication shall be provided for monitoring of AC supply healthiness through DC operated fascia annunciation of bus coupler panel.

PT Secondary Supply Distributions
      The bus coupler panel shall receive PT secondary volt through fuses and link provided in PT kiosk in the switchyard. Selected PT secondary supply to protective relays of each panel shall be fed through individual fuses and links. Necessary arrangement for supervision of PT secondary supply is provided.Selected PT secondary supply to metering and indicating instrument of each panel shall be fed through fuses and links.

Annunciation Scheme

Other trip and non trip alarm scheme. Fascia annunciation system has to be provided in each control panel by means of visual and audible alarm to draw attention of operator. The annunciation can be divided into the following categories
Trip annunciation
Warning annunciation

      Incoming DC fail and annunciation bus DC fail alarm scheme. Bus coupler panel shall have an "incoming DC fail and annunciation bus DC fail alarm scheme" common to the entire C&R board and operate from 230V   AC supply for audible and visual alarm through lamp and bell arrangement. The scheme comprises of DC supervision relay with push buttons for incoming DC supply and annunciation bus DC system, one AC operated alarm accept relays, one indication lamp, one AC operated hooter and one push button for cancellation of audible alarm.Alarm inscriptions are engraved in facial window and the same should be prominently visible when fascia light is ON.Accept reset and lamp test push buttons for acknowledgement of alarm, reset of visual indicator and checking correctness of lamps are provided below each fascia.The annunciation scheme should be such that momentary closing of any fault contacts shall cause operation of annunciation.

Trip Annunciation
      Trip annunciation shall draw attention of operation when a C.B. is tripped automatically thorough relay. The audible common alarm should buzz and the visual annunciation is provided by flickering of window fascia.

Warning Annunciation
      Warning annunciation is used to draw attention of operator to abnormal operating conditions of equipment for immediate action to avoid tripping of C.B. audible alarm is provided by a separate common bell with different tone and visual annunciation of flickering of fascia window.

Protection Philosophy
      The philosophy of protection is to design the protection scheme for lines, transformers, bus couplers, bus transfer bay depending upon the substation layout arrangement. The scheme shall be generally based on the requirement listed below.

Line Protections
 Numerical distance relay with built-in auto enclosure or in standalone mode.
Separate fault locator or included in distance relay (available in ABB make RES SII type relay).
 Disturbance Recorder (may be built in with Distance Relay).
Back up directional over-current relay
Back up directional earth fault relay.

132/33KV Autotransformer Protection
Percentage biased differential protection(3winding)
Backup directional voltage potential 3element IDMT over current with high set instantaneous element
Backup directional voltage polarized earth fault relay with high set instantaneous element.
Local break up back up protection at HV side only.
Over flux protection and overload alarm at HV side.
Flag relays for internal protection of transformer like buchholz relay, winding temperature, oil temperature and pressure relief Valve

132 KV Bus Coupler Protections
Bus differential protection scheme.
 Local breaker back up protection.
AC supervision relay.

DC supervision relay.
 DC fail relay
3phase trip relay.

132KV Transfer Bus Bay Protections
 The breaker is used for transfer of any circuit through transfer bus in case of necessity. During the period of transfer the line side CT operated protection of transferred circuit will trip the transfer bus breaker in place of respective bay circuit breaker through trip transfer switch. Normally the protection scheme provided in transfer bus bay excepting the LBB protection shall be kept in operating with the help of protection IN/OUT selector switch. However, its own protection shall be used when there is trouble in the protection system of transferred circuit.

The protection scheme of this bay will be as follows
Numerical distance protection scheme for transferred feeder protection with protection IN/OUT switch but without carrier intertripping, auto reclose relay, in-built fault locater, disturbance recorder.
Local breaker backup protection.
Backup directional IDMT overcurrent relay with protection IN/OUT switch.
Backup directional earth fault relay.

Clamps and Connectors
 The role of clamps and connectors can be felt during operation of substation. Connectors of bad quality often result in outage of system and can even lead to total bus fault. The cause of failure of connector is mainly due to heating and excessive electromechanical stress. That is why, choice of materials and design of connectors are given due consideration while designing the substation.

Isolators
 When carrying out inspection or repair in a substation installation, it is essential to disconnect reliably the unit or section on which the work is to be done, from all other live parts on the installation in order to ensure complete safety of the working staff. To guard against mistakes it is desirable that this is done by an apparatus which makes a visible break in the circuit. Such an apparatus is the isolating switch or ISOLATOR.Isolators used in power systems are generally three pole isolator. The three pole isolators have three identical poles. Each pole consists of two or three isolator posts mounted on fabricated supports. The conducting parts are supported on insulator posts. The conducting part consists of conducting copper or aluminum rod, fixed and moving contacts. During the opening operation the conducting rod swings apart and isolation is obtained.
The simultaneous operation of 3 poles is obtained by mechanical interlocking of 3 poles. For all 3 poles there is a common operating mechanism. The operating mechanism is manual plus one of the following
- Pneumatic mechanism
- Electric motor mechanism

 Pneumatic mechanism is the primitive mechanism of isolator operation. It operates on compressed air at 8 kg/cm^2, example of this type of isolator is a TMG make isolator.But the biggest disadvantage of pneumatic operation is that the whole process is dependent on compressor plant station. In case of breakdown in compressor plant or the airline, isolator becomes dysfunctional.Now days, however motor driven isolators are put to use. These are operated by 3φ ac motors. In case of breakdown of electrical system, manual operation is also possible.The pneumatic operated isolators are used only where a source of compressed air is available. Motorized systems are usually simpler and less costly and hence we are using these in our design.The operating mechanism is generally mounted direct on base frame of the isolator. The actuator unit then requires a bearing and additional link rod when the system is well off the ground. Emergency manual operation is possible with all operating mechanisms, if the power fails.The operating mechanisms also incorporate actuating switches for indicating the switching position and for control and interlock purposes. Motor driven units also include contactors and control devices. The control system is arranged so that only one switching pulse is needed and the actuators switch off automatically when end position is reached. In the event of emergency manual operation safety contact interrupts the motor circuit so that simultaneous actuation from control room is not possible.
To prevent maloperation, the operating mechanism of isolators and earthing switches can be interlocked relative to each other, motorized systems electrically, compressed air systems electro-pneumatically and manual systems mechanically.

IX.   CONCLUSION

In conclusion to all the mentioned design aspects of the 132/33KV sub-station there are several other factors that are needed to be considered. This includes socio-economic factor of the surrounding locality, political developments, union of workers and contractors. Economic factors become chief aspect in any project which can take a prolonged period to complete. An assumption of price hike of all the materials to a higher precision is needed to be made in order to estimate the budget of this project.The mechanical and civil designs are also an essential part of any electrical substation design. Thus a lot of other engineering brains in those fields are also employed for the construction. Experts in the field of commerce and law are also employed to meet the various challenges that may rise up. It's an overall build up that ensures huge employment of people from different fields.

REFERENCES

Books:
[1]   A textbook of Electrical Power System by V.K.Mehta.
[2]   ELECTRIC POWER SUBSTATIONS ENGINEERING, THIRD EDITION (ELECTRICAL ENGINEERING HANDBOOK) BY JOHN D. MCDONALD
[3]   PROTECTIVE RELAYING: PRINCIPLES AND APPLICATIONS, THIRD EDITION (POWER ENGINEERING (WILLIS)) BY J. LEWES BLACKBURN
[4]   POWER SYSTEM RELAYING (RSP) BY STANLEY H. HOROWITZ
[5]   POWER SYSTEM PROTECTION (IEEE PRESS SERIES ON POWER ENGINEERING) BY PAUL M. ANDERSON
[6]   ELECTRICAL POWER TRANSMISSION SYSTEM ENGINEERING: ANALYSIS AND DESIGN, 2ND EDITION BY TURAN GONEN
[7]   PROTECTION OF ELECTRICITY DISTRIBUTION NETWORKS, 3RD EDITION (IET POWER AND ENERGY) BY JUAN GERS AND EDWARD J. HOLMES
[8]   Power Circuit Breaker Theory and Design by Charles H. Flurscheim
[9]   Protection and Switchgear by U.A.Bakshi, M.V.Bakshi
[10]   Conductors and Insulators by Angela Royston

Noise Supression with Triple Phase Sllep Signal Slew Rate Modulation in Mtcmos Circuits with Power Gating Methods

[1,]T.Suhasini , [2,] G.Nagajyothi

[1,]*M.Tech(Vlsi),SSITS,Rayachoti,Kadapa Dist,India*
[2,]*Assistant Prof,SSITS,Rayachoti,Kadapa Dist,India*

ABSTRACT:
As low power circuits are most popular now a days as the scaling increase the leakage power in the circuit also increases rapidly so for removing these kind of leakages and to provide a better power efficiency we are using many types of power gating techniques. In this paper we are going to analyse the different types of flip-flops using different types of power gated circuits using low power VLSI design techniques and we are going to display the comparison results between different nanometer technologies. The NMOS simulations were done using Microwind Layout Editor & DSCH software and the results were given below.

I. INTRODUCTION

The scaling of process technologies to nanometer regime has resulted in a rapid increase in leakage power disNMOS1pation. Hence, it has become extremely important to develop design techniques to reduce static power disNMOS1pation during periods of inactivity. The power reduction must be achieved without trading-off performance which makeNMOS1t harder to reduce leakage during normal (runtime) operation. On the other hand, there are several techniques for reducing leakage power in *sleep* or *standby* mode. Power gating is one such well known technique where a *sleep transistor* is added between actual ground rail and circuit ground (called *virtual ground*). This device is turned-off in the sleep mode to cut-off the leakage path. It has been shown that this technique provides a substantial reductionin leakage at a minimal impact on performance

Power gating technique uses high Vt sleep transistors which cut off VDD from a circuit block when the block is not switching. The sleep transistor NMOS1zing is an important design parameter. This technique, also known as MTCMOS, or Multi-Threshold CMOS reduces stand-by or leakage power, and also enableNMOS1ddq testing.

1.1.Power gating: affects design architecture more than clock gating. It increases time delays as power gated modes have to be safely entered and exited. Architectural trade-offs exist between designing forthe amount of leakage power saving in low power modes and the energy disNMOS1pation to enter and exit the low power modes. Shutting down the blocks can be accomplished either by software or hardware. Driver software can schedule the power down operations. Hardware timers can be utilized. A dedicated power management controller is another option.An externally switched power supply is a very baNMOS1c form of power gating to achieve long term leakage power reduction. To shut off the block for small intervals of time, internal power gating is more suitable. CMOS switches that provide power to the circuitry are controlled by power gating controllers. Outputs of the power gated block discharge slowly. Hence output voltage levels spend more time in threshold voltage level.

This can lead to larger short circuit current.Power gating uses low-leakage PMOS transistors as header switches to shut off power supplies toparts of a design in standby or sleep mode. NMOS footer switches can also be used as sleep transistors. Inserting the sleep transistors splits the chip's power network into a permanent power network connected to the power supply and a virtual power network that drives the cells and can be turned off.The quality of this complex power network is critical to the success of a power-gating design. Two of the most critical parameters are the IR-drop and the penaltieNMOS1n NMOS1licon area and routing resources. Power gating can be implemented using cell- or cluster-based (or fine grain) approaches or a distributed coarse-grained approach.
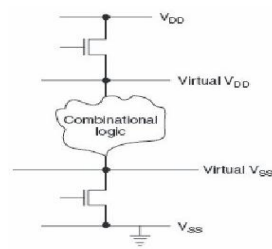
Fig1: Power Gated Circuits

II. POWER-GATING PARAMETERS

Power gating implementation has additional conNMOS1derations for timing closure implementation. The following parameters need to be conNMOS1dered and their values carefully chosen for a successful implementation of this methodology.

[1] *Power gate Size*: The power gate Size must be selected to handle the amount of switching current at any given time. The gate must be bigger such that there is no measurable voltage (IR) drop due to the gate. As a rule of thumb, the gate Size is selected to be around 3 times the switching capacitance. Designers can also choose between header (P-MOS) or footer (N-MOS) gate. Usually footer gates tend to be smaller in area for the same switching current. Dynamic power analyNMOS1s tools can accurately measure the switching current and also predict the Size for the power gate.

[2] *Gate control slew rate*: In power gating, thiNMOS1s an important parameter that determines the power gating efficiency. When the slew rate is large, it takes more time to switch off and switch-on the circuit and hence can affect the power gating efficiency. Slew rate is controlled through buffering the gate control Signal.

[3] *NMOS1multaneous switching capacitance*: ThiNMOS1mportant constraint refers to the amount of circuit that can be switched NMOS1multaneously without affecting the power network integrity. If a large amount of the circuit is switched NMOS1multaneously, the resulting "rush current" can compromise the power network integrity. The circuit needs to be switched in stageNMOS1n order to prevent this.

[4] *Power gate leakage*: NMOS1nce power gates are made of active transistors, leakage reduction is an important conNMOS1deration to maximize power savings.

2.1.Fine-grain power gating

Adding a sleep transistor to every cell that is to be turned off imposes a large area penalty, and individually gating the power of every cluster of cells creates timing issueNMOS1ntroduced by inter-cluster voltage variation that are difficult to resolve. Fine-grain power gating encapsulates the switching transistor as a part of the standard cell logic. Switching transistors are designed by either the library IP vendor or standard cell designer. Usually these cell designs conform to the normal standard cell rules and can eaNMOS1ly be handled by EDA tools for implementation.The Size of the gate control is designed conNMOS1dering the worst case scenario that will require the circuit to switch during every clock cycle, resulting in a huge area impact. Some of the recent designNMOS1mplement the fine-grain power gating selectively, but only for the low Vt cells. If the technology allows multiple Vt libraries, the use of low Vt deviceNMOS1s minimum in the design (20%), so that the area impact can be reduced. When using power gates on the low Vt cells the output must be isolated if the next stage is a high Vt cell. Otherwise it can cause the neighboring high Vt cell to have leakage when output goes to an unknown state due to power gating.

Gate control slew rate constraint is achieved by having a buffer distribution tree for the control Signals. The buffers must be chosen from a set of always on buffers (buffers without the gate control Signal) designed with high Vt cells. The inherent difference between when a cell switches off with respect to another, minimizes the rush current during switch-on and switch-off.

Usually the gating transistor is designed as a high Vt device. Coarse-grain power gating offers further flexibility by optimizing the power gating cells where there is low switching activity. Leakage optimization has to be done at the coarse grain level, swapping the low leakage cell for the high leakage one. Fine-grain power gating is an elegant methodology resulting in up to 10 times leakage reduction. This type of power reduction

makeNMOS1t an appealing technique if the power reduction requirement is not satisfied by multiple Vt optimization alone.

2.2.Coarse-grain power gating

The coarse-grained approach implements the grid style sleep transistors which drives cells locally through shared virtual power networks. This approach is less senNMOS1tive to PVT variation, introduces lesNMOS1R-drop variation, and imposes a smaller area overhead than the cell- or cluster-based implementations. In coarse-grain power gating, the power-gating transistor is a part of the power distribution network rather than the standard cell.

There are two ways of implementing a coarse-grain structure:
[1]   *Ring-based*: The power gates are placed around the perimeter of the module that is being switched-off as a ring. Special corner cells are used to turn the power Signals around the corners.
[2]   *Column-based*: The power gates are inserted within the module with the cells abutted to each other in the form of columns. The global power is the higher layers of metal, while the switched power iNMOS1n the lower layers.

Gate NMOS1zing depends on the overall switching current of the module at any given time. NMOS1nce only a fraction of circuits switch at any point of time, power gate Sizes are smaller as compared to the fine-grain switches. Dynamic power NMOS1mulation using worst case vectors can determine the worst case switching for the module and hence the Size. The IR drop can also be factored into the analyNMOS1s. NMOS1multaneous switching capacitance is a major conNMOS1deration in coarse-grain power gating implementation. In order to limit NMOS1multaneous switching, gate control buffers can be daisy chained, and special counters can be used to selectively turn on blocks of switches.

III.   POWER GATING FOR DELAY REDUCTION
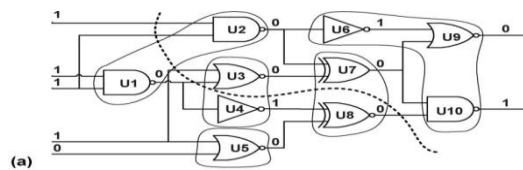
Fig2: Device without Power gating.
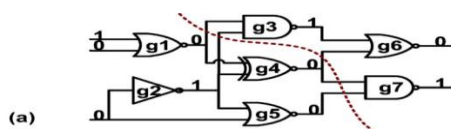
Fig3: Device with Power gating with reduced area & Power using clustering network formation.

This work presented a logic clustering based solution to the problem of controlling/optimizing the power gating parameters. The key design conNMOS1derationNMOS1n the power mode transitions are minimizing the wakeup delay, the peak current, and the total Size of sleep transistors. This work analyzed the relations between the three parameters, and solved the problem of finding logic clusters and their wakeup schedule that minimize the wakeup delay while satisfying the peak current and performance loss constraints.
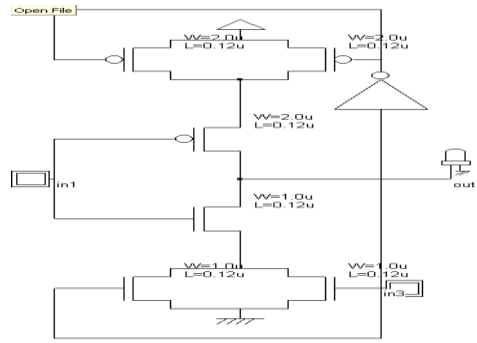
Fig4: Sleepy stack
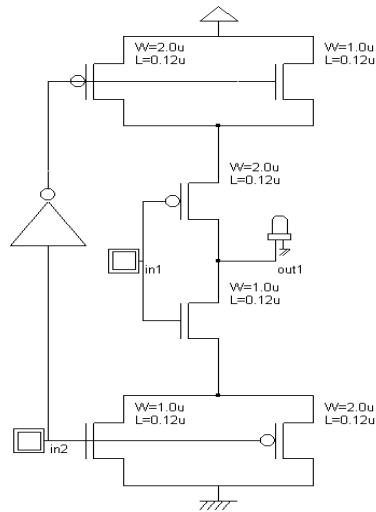
Fig5: Dual Sleep Method
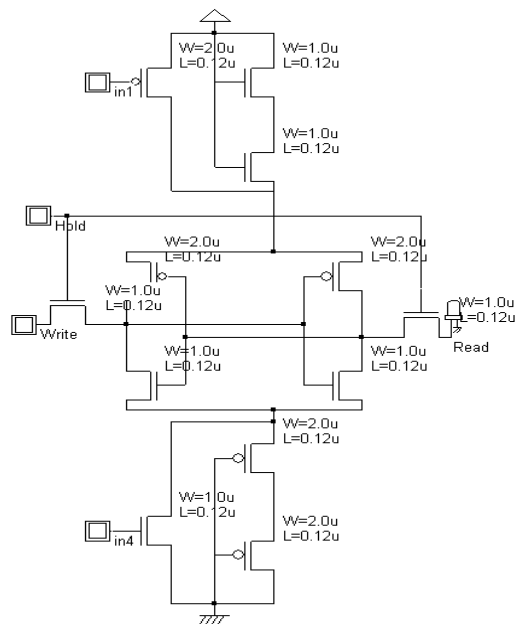
Fig6: Dual Stack Approach

A variation of the sleep approach, the zigzag approach, reduces wake- up overhead caused by sleep

transistors by placement of alternating sleep transistors assuming a particular pre-selected input vector [6]. Another technique for leakage power reduction is the stack approach, which forces a stack effect by breaking down an existing transistor into two half Size transistors [7]. The divided transistorNMOS1ncrease delay Significantly and could limit the usefulness of the approach. The sleepy stack approach (Fig. 2) combines the sleep and stack approaches [2, 3]. The sleepy stack technique divides existing transistor NMOS1nto two half Size transistors like the stack approach. Then sleep transistors are added in parallel to one of the divided transistors. During sleep mode, sleep transistors are turned off and stacked transistors suppress leakage current while saving state. Each sleep transistor, placed in parallel to the one of the stacked transistors, reduces resistance of the path, so delay is decreased during active mode. However, area penalty is a NMOS1gnificant matter for this approach NMOS1nce every transistor is replaced by three transistors and NMOS1nce additional wires are added for S and S', which are sleep Signals. Another technique called Dual sleep approach [8] (Fig. 3) uses the advantage of using the two extra pull- up and two extra pull-down transistorNMOS1n sleep mode either in OFF state or in ON state. NMOS1nce the dual sleep portion can be made common to all logic circuitry, less number of transistorNMOS1s needed to apply a certain logic circuit.

IV.   PROPOSED POWER GATING BASED SLEEP TECHNIQUE
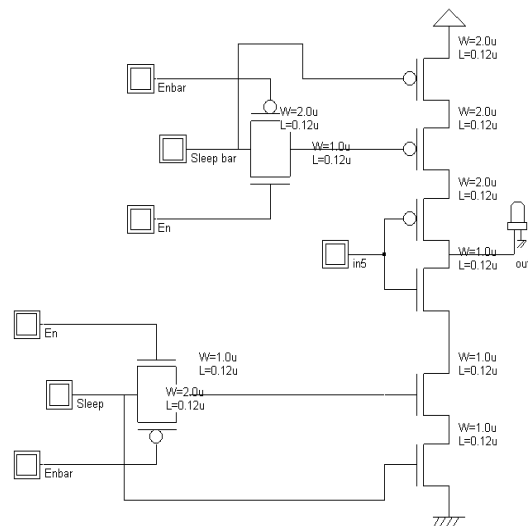
The Above mentioned Sleep Circuit has three modes of operations

[1]   Active mode
[2]   Standby mode
[3]   Sleep to active mode transition

In active mode, the sleep Signal of the transistor is held at logic' 1 ' and both the sleep transistors **M I and M2 (En and EnBar Transistors from the bottom side)** remain ON. In this case both transistors offer very low resistance and virtual ground (VGND) node potential is pulled down to the ground potential, making the logic difference between the logic circuitry approximately equal to the supply voltage.

There are several benefits of combining stacked sleep transistors. First the magnitude of power supply fluctuations sleep mode during mode transitions will be reduced because these transitions are gradual. Second, while conventional power gating uses a high- threshold device as a sleep transistor to minimize leakage, a stacked sleep structures can achieve the same effect with a normal threshold device

In active mode, the sleep Signal of the transistor is held at logic' 1 ' and both the sleep transistors **NMOS1 and NMOS2 (NMOS Transistors used for sleep Purpose from the Bottom of the circuit)** remain ON and control transistor is OFF by giving logic O. In this case both transistors offer very low resistance and virtual ground (VGND) node potential is pulled down to the ground potential, making the logic difference between the logic circuitry approximately equal to the supply Voltage. And leakage current is reduced by the stacking effect, turning both NMOS1 and NMOS2 sleep transistors OFF. **And vice versa for the header switch.**

Positive potential at the intermediate node has four effects:

- Gate to source voltage of NMOS1 (VgNMOS1) becomes negative.
- -Negative body -to-source potential (Vdsl) of NMOS1 decreases, resulting in less drain induced barrier lowering.
- -Drain-to-source potential (VdNMOS2) of NMOS2 is less compared to NMOS1, because most of the voltage drops across the NMOS1 in sleep mode.

This Significantly reduces the drain barrier lowering. The analyzed design gives major contribution in sleep to active mode in terms of peak of sleep mode compared to stacking power gating. Sleep mode occurs when circuit is going from sleep to active and vice versa. In first stage sleep transistor (NMOS1) working as diode by turn on the control transistor M I which is connected across the drain and gate of the sleep transistor (NMOS1). Due to this drain to source current of the sleep transistor dropNMOS1n a quadratic manner. This reduces the voltage fluctuation on the ground and power net and it also reduces the circuit wakeup time. So in sleep to active transition mode, we are turning ON transistor NMOS1 initially after small duration of time NMOS2 will be turned ON to reduce the GBN. In second stage control transistor is off that sleep transistor works normally.During sleep to active mode transition, transistor NMOS1 is turned ON and transistor NMOS2 is turned ON after a small duration of time (6 T). The logic circuit iNMOS1solated from the ground for a short duration as the transistor NMOS2 is turned OFF. During this duration, the GBN can be greatly reduced by controlling the intermediate node voltage VGND2 and operating the transistor NMOS2 in triode region.The intermediate node (VGND2) voltage can by Inserting proper amount of delay, that is less than the discharging time of the NMOS1 transistor.

Proper selection of the capacitance C2. Leakage current is reduced by the stacking effect, turning both NMOS1 and NMOS2 sleep transistors OFF. This raises the intermediate node voltage VGND2 to positive values due to small drain current. Positive potential at the intermediate node has four effects: Gate to source voltage of NMOS1 (VgNMOS1) becomes negative. Negative body- to- source potential (VbNMOS1) of NMOS1 causes more body effect Drain- to- source potential (Vdsl) of NMOS1 decreases, resulting in less drain induced barrier lowering.Drain-to-source potential (VdNMOS2) of NMOS2 is less compared to NMOS1, because most of the voltage drops across the NMOS1 in sleep mode this Significantly reduces the drain induced barrier lowering.

V. CONCLUSION

In nanometer scale CMOS technology, sub threshold leakage power consumption is a great challenge. Although previous approaches are effective in some ways, no perfect solution for reducing leakage power consumption is yet known. Therefore, designers choose techniques based upon technology and design criteria. In this paper, we provide novel circuit structure named "Dual stack" as a new remedy for designer in terms of static power and dynamic powers. Unlike the sleep transistor technique, the dual stack technique retains the original state. The dual stack approach shows the least speed power product among all methods. Therefore, the dual stack technique provides new ways to designers who require ultra-low leakage power consumption with much less speed power product. Especially it shows nearly 50-60% of power than the existing normal or conventional flip-flops. So, it can be used for future integrated circuits for power & area Efficiency.

REFERENCES

[1] M. Powell, S.-H. Yang, B. Falsafi, K. Roy and T. N. Vijaykumar, "Gated-Vdd: A Circuit Technique to Reduce Leakage in Deep submicron Cache Memories," *Proc. of International SympoNMOS1um onLow Power Electronics and Design*, pp. 90-95, July 2000.

[2] J.C. Park, V. J. Mooney III and P. Pfeiffenberger, "Sleepy Stack Reduction of Leakage Power," *Proc. of the International Workshop onPower and Timing Modeling, Optimization and NMOS1mulation*, pp. 148-158, September 2004.

[3] J. Park, "Sleepy Stack: a New Approach to Low Power VLSI and Memory," Ph.D. Dissertation, School of Electrical and Computer Engineering, Georgia Institute of Technology, 2005. [Online].Available http://etd.gatech.edu/theses.

[4] S. Mutoh, T. Douseki,. Y. Matsuya, T. Aoki, S. Shigematsu and J. Yamada, "1-V Power Supply High-speed Digital Circuit Technology with Multithreshold-Voltage CMOS," *IEEE Journal of Solis-StateCircuits*, vol. 30, no. 8, pp. 847–854, August 1995.

[5] N. Kim, T. Austin, D. Baauw, T. Mudge, K. Flautner, J. Hu, M. Irwin, M. Kandemir and V. Narayanan, "Leakage Current: Moore's Law Meets Static Power," *IEEE Computer*, vol. 36, pp. 68–75, December 2003.

[6] K.-S. Min, H. Kawaguchi and T. Sakurai, "Zigzag Super Cut-off CMOS (ZSCCMOS) Block Activation with Self-Adaptive Voltage Level Controller: An Alternative to Clock-gating Scheme in Leakage

[7] J. Shin and T. Kim, "Technique for transition energy-aware dynamicvoltage asNMOS1gnment," *IEEE Trans. Integr. Circuits Syst. II, Exp. Briefs*,vol. 53, no. 9, pp. 956–960, Sep. 2006.

[8] W. Cheol and T. Kim, "Optimal voltage allocation techniques fordynamically variable voltage processors," *ACM Trans. EmbeddedComput. Syst.*, vol. 4, no. 1, pp. 211–230, Feb. 2005.

[9] T. Ishihara and H. Yasuura, "Voltage scheduling problem for dynamically variable voltage processors," in *Proc. IEEE/ACM Int. Symp. LowPower Electron. Des.*, 1998, pp. 197–202.

[10] F. Fallah and M. Pedram, "Standby and active leakage current controland minimization CMOS VLSI circuits," *IEICE Trans. Electron.*, vol.E88-C, no. 4, pp. 509–519, 2005.

[11] J. Friedrich, B. McCredie, N. James, B. Huott, B. Curran, E. Fluhr, G.Mittal, E. Chan, Y. Chan, D. Plass, S. Chu, H. Le, L. Clark, J. Ripley, S.Taylor, J. Dilullo, and M. Lanzerotti, "Design of the Power6 microprocessor,"in*Proc. IEEE/ACM Int. Solid-State Circuits Conf.*, Feb. 2007,pp. 96–97.

[12] S. Mutoh, T. Douseki, Y. Matsuya, T. Aoki, S. Shigematsu, and J. Yamada,"1-V power supply high-speed digital circuit technology with multi-threshold voltage CMOS," *IEEE J. Solid-State Circuits*, vol. 30, no. 8, pp. 847–854, Aug. 1995.

[13] J. Kao, A. Chandrakasan, and D. Antoniadis, "Transistor NMOS1zing issuesand tool for multi-threshold CMOS technology," in *Proc. IEEE/ACMDes. Autom. Conf.*, 1997, pp. 409–414.

[14] D. Chiou, S. Chen, S. Chang, and C. Yeh, "Timing driven powergating," in *Proc. IEEE/ACM Des. Autom. Conf.*, 2006, pp. 121–124.

[15] A. Sathanur, L. Benini, A. Macii, E. Macii, and M. Poncion, "Multiplepower-gating domain(multi-vgnd) architecture for improved leakagepower reduction," in *Proc. IEEE/ACM Int. Symp. Low Power Electron.Des.*, 2008, pp. 51–56.

[16] F. Li and L. He, "Maximum current estimation conNMOS1dering powergating," in *Proc. IEEE/ACM Int. Symp. Low Power Electron. Des.*,2001, pp. 409–414.

[17] H. Jiang and M. Marek-Sadowska, "Power gating scheduling forpower/ground noise reduction," in *Proc. IEEE/ACM Des. Autom.Conf.*, 2008, pp. 980–985.

[18] S. Kim, S. Kosonocky, and D. Knebel, "Understanding and minimizingground bounce during mode transition of power gating structures,"in*Proc. IEEE/ACM Int. Symp. Low Power Electron. Des.*, 2003, pp.22–25.

[19] Y. Chen, D. Juan, M. Lee, and S. Chang, "An efficient wake-upschedule during power mode transition conNMOS1dering spurious glitchesphenomenon," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Des.*2007, pp. 779–782.

[20] C. Long and L. He, "Distributed sleep transistor network for powerreduction," in *Proc. IEEE/ACM Des. Autom. Conf.*, 2003, pp. 181–187.

[21] A. Abdollahi, F. Fallah, and M. Pedram, "An effective power modetransition technique in MTCMOS circuits," in *Proc. IEEE/ACM Des. Autom. Conf.*, 2005, pp. 37–42.

[22]

Power Quality Improvement Using Ac To Ac PWM Converter for Distribution Line

Shalini Bajpai

Jabalpur Engineering Collage Jabalpur MadhyaPradesh

ABSTRACT:

In this paper, a new voltage sag compensator for critical loads in electrical distribution system discussed. The proposed scheme employs a Pulsewidth modulation ac-ac convertor along with an auto transformer. During a disturbance such as voltage sag, the proposed scheme supplies the missing voltage and helps in maintaining rated voltage at the terminals of the critical load. Under normal condition the approach works in bypass mode and delivering utility power directly to load. A four step switching technique ti drive the ac-ac convertor is employed to realize snubberless operation. A design is presented for 440v, 50 hz, system.

KEYWORDS:- *Power Quality, Voltage Sag, PWM Ac-Ac Converter*

I.   INTRODUCTION: POWER QUALITY

POWER QUALITY is a term that mean different to different people. Institute of Electrical and electrical engineers (IEEE) standard IEEE 1100 defines power quality "as the concept of powering and grounding sensitive electronic equipment in a manner suitable for the equipment". In a simpler words PQ is a set of electrical boundaries that allows a part of equipment to work in a intended manner without loss of performance or life expectancy.

In recent years power engineers are increasingly concerned over PQ due to following reasons:

1.1.REASONS FOR INCREASED CONCERN REGARDING POWER QUALITY

The most responsible reason is the newer-generation load equipment with microprocessor-based controls and power electronic devices such loads are more sensitive to power quality variations than was equipment used in the past.The another reason is the increased use of power electronics devices such as electrical drives ,fact devices, static relays etc due to the increased emphasis on improving overall power system efficiency . This is resulting in increasing harmonic levels on power systems and has many people oncerned about the future impact on system capabilities.Increased awareness of power quality issues among the end users.Most of the systems are now interconnected in a network. Hence the processes are integrated in which the failure of anycomponent can results into important consequences.The globalization of industries has increased the awareness about deficiencies in power quality around the world.The economic value of power quality is also one important reason for its increased concern. There is a big money associated with these power quality disturbances. Many efforts have been taken by the utilities to meet the consumer's PQ requirements. Hence FACT devices and various other custom power devices are introduced in the electrical system to improve the PQ of the electrical power.

1.2.TYPES OF POWER QUALITY PROBLEMS

Voltage sag (or dip)

It is a dip of .1 to .9 p.u. in rms voltage or current at the power frequency, for interval of 0,5 cycle to 1 minute.

Causes: Whenever a load end draws a heavy current suddenly. That's why it is associated with faults on the transmission or distribution network, faults in consumer's installation , sudden connection of heavy loads and start-up of large motors.

Consequences: Malfunction of microprocessor-based control systems (PCs, PLCs, ASDs, etc), that may cause false tripping of contactors and electromechanical relays. Maloperation of electric rotating machines.
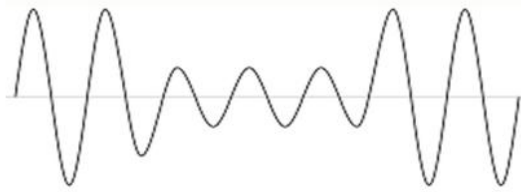
Fig.1-Voltage sag

II.   VERY SHORT INTERRUPTIONS

It is a total interruption or decrease of supply voltage or load current to less than .1 p.u. for few milliseconds to one or two seconds.

Causes: It occurs due to failure of protecting devices, insulation failure, control malfunction. Also due to delayed reclosing of the protecting devices.

Consequences: False tripping of protection devices, loss of information results in malfunction of data processing equipment. Sensitive equipment stops working.
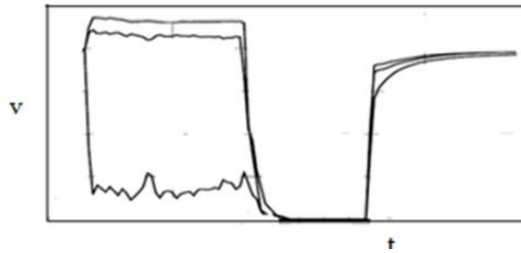
Fig.2- Very short interruptions

III.   LONG INTERRUPTIONS

Total interruption of electrical supply for duration greater than 1 to 2 seconds

Causes: Equipment failure in the power system network, storms and objects (trees, cars, etc) striking lines or poles, fire, human error, bad coordination or failure of protection devices.

Consequences: Stoppage of all equipment.
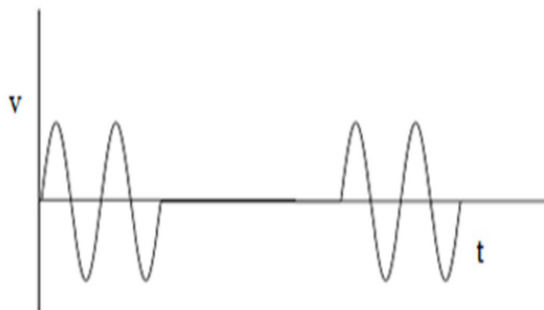
Fig.3- Long interruptions

IV.   VOLTAGE SPIKE

It is the very fast variation of the voltage for durations from a several microseconds to few milliseconds. These variations may reach thousands of volts, even in low voltage.

Causes: Lightning which is a natural cause , switching of lines or power factor correction capacitors ,sudden removal of heavy loads.

Consequences: Damage of components (particularly electronic components) and of insulation materials, data processing errors , electromagnetic interference or information loss.

Fig.4- Voltage Spike

V.   VOLTAGE SWELL

Momentary increase of the voltage, at the power frequency, outside the normal tolerances, with duration of more than one cycle and typically less than a few seconds.

Causes: Start/stop of heavy loads, badly dimensioned power sources, badly regulated transformers (mainly during off-peak hours).

Consequences: Data loss, flickering of lighting and screens, stoppage or damage of sensitive equipment, if the voltage values are too high.
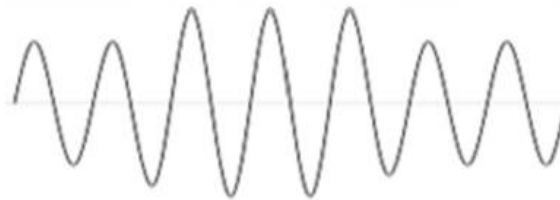
Fig.5- Voltage swell

VI.   HARMONIC DISTORTION

These are periodically distorted voltage or current waveform . The waveform corresponds to the sum of different sine-waves with different magnitude and phase, having frequencies that are multiples of power-system frequency.

Causes: Arc furnances electric machines working above the magnetic saturation, welding machines, rectifiers, and DC brush motors. All non-linear loads, such as power electronics equipment including ASDs, switched mode power supplies, high efficiency lighting, data processing equipments.

Consequences: Probability of occurrence severe resonance increases, neutral overloading in 3-phase systems, overheating of all cables and equipment, loss of efficiency in electric machines, electromagnetic interference with communication systems, errors in measures when using average reading meters, nuisance tripping of thermal protections ,can induce visual flicker in arc lighting.
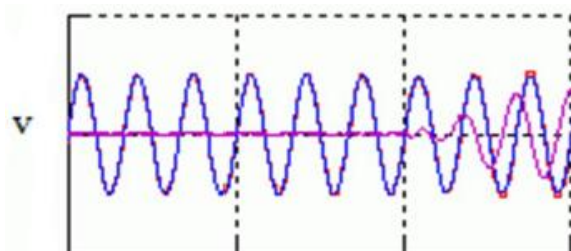
Fig.-6 Harmonic distortion

VII. VOLTAGE FLUCTUATION

It is a series of random voltage variations or systematic variations of voltage envelop but the variation does not exceeds the voltage ranges of 0.9 to 1.1 p.u. . Oscillation of voltage value, amplitude modulated by a signal with frequency of 0 to 30 Hz.

Causes: Frequent start/stop of electric motors (for instance elevators), oscillating loads, arc furnances

Consequences: . The most perceptible consequence is the flickering of lighting and screens, giving the impression of unsteadiness of visual perception and the rest effects are similar to undervoltages.
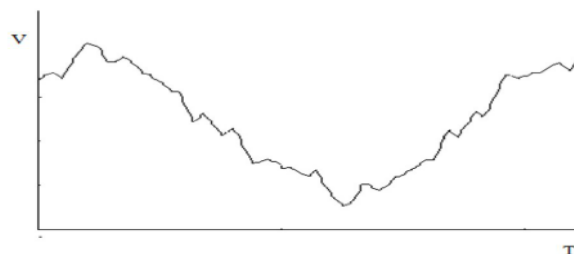
Fig.-7 Voltage fluctuation

VIII. NOISE

It is the Superimposing of high frequency unwanted signals on the waveform of the power-system frequency.

Causes: Electromagnetic interferences provoked by microwaves, television diffusion, and radiation due to welding machines, arc furnaces, and electronic equipment. Improper grounding may also be a cause.

Consequences: Disturbances on sensitive electronic equipment, usually not destructive. May cause some data related errors .

IX. VOLTAGE UNBALANCE

A maximum voltage variation in a three-phase system in which the three voltage magnitudes or the phase angle differences between them are not equal.

Causes:, Incorrect distribution of all single-phase loads by the three phases of the system (this may be also due to a fault). Large single-phase loads (induction furnaces, traction loads)

Consequences: It mostly affects three-phase induction machines. Unbalanced systems imply the existence of a negative sequence that is harmful to all three phase loads.

It has found that among all the PQ issues voltage sag and swell are the most occurring problems at the distribution end [1]-[2]-[3],as shown in fig.8-
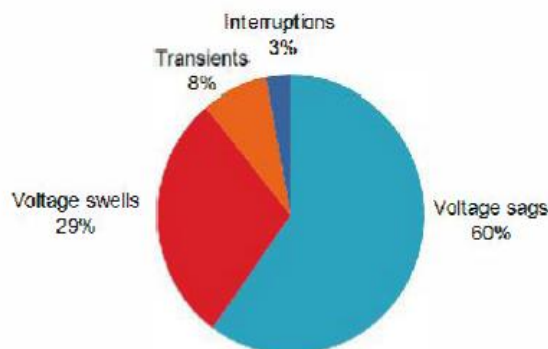
Fig.-8 Showing the occurring percentage of various PQ issues.

From fig.8 we can see that the voltage sag swell issues cover almost 89% of the graph area. That means they are the most responsible issues for decreasing power quality.

.

Proposed system configuration

The proposed device for mitigating voltage sag and swell in the system consists of a PWM switched power electronic device connected to an autotransformer in series with the load.
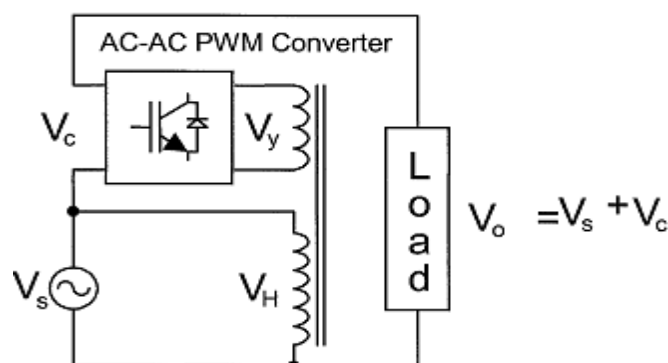
Fig. 9 shows the single phase circuit configuration of the mitigating device and the control circuit logic used in the system. It consists of a single PWM insulated gate bipolar transistor (IGBT) switch in a bridge configuration, a thyristor bypass switch, an autotransformer, and voltage controller.

PRINCIPLE OF OPERATION

To maintain the load voltage constant an IGBT is used as power electronic device to inject the errorvoltage into the line. Four power diodes (D1 to D4) connected to IGBT switch (SW) controls the direction of power flow and connected in ac voltage controller configuration with a suitable control circuit maintains constant rms load voltage. In this scheme sinusoidal PWM pulse technique is used. RMS value of the load voltage VL is calculated and compared with the reference rms voltage Vref. During normal condition the power flow is through the anti parallel thyristors. Output filters containing a main capacitor filter and a notch filter are used at the output side to filter out the switching noise and reduce harmonics. During this normal condition, VL = Vref and the error voltage Verr is zero. The gate pulses are blocked to IGBT. Due to sudden increase or decrease in the load or due to faults voltage sag or swell occurs. The supply voltage VS and hence VL decreases. When the sensing circuit detects an error voltage Verr greater than 10% of the normal voltage the voltage controller acts immediately to switch off the thyristors. Voltage Verr applied tothe pi controller gives the phase angle delta .The phase angle delta is dependent on the percentage of disturbance and hence controls the magnitude of Vcontrol. This control voltage is then compared with the triangular voltage Vtri to generate the PWM pulses VG which are applied to the IGBT to regulate the output voltage. Hence the IGBT switch operates only during voltage sag or swells condition and regulates the output voltage according to the PWM duty-cycle. To suppress the over voltage when the switches are turned off, RC snubber circuits are connected across the IGBT and thyristor.

SIMULATION ANALYSIS AND RESULTS

Simulation analysis is performed on a three-phase, 115/11 kV, 100 MVA, 50 Hz system to study the performance of the PWM switched autotransformer in mitigating the voltage sag and swell disturbances. The MATLAB/SIMULINK model of the system used for analysis is shown An RL load is considered as a sensitive load, which is to be supplied at constant voltage. Tab. 1 shows the system parameter specifications used for simulation. Under normal condition, the power flow is through the antiparallel SCRs and the gate pulses are inhibitedto IGBT. The load voltage and current are same as supply voltage and current. When a disturbance occurs, an error voltage which is the difference between the reference rms voltage and the load rms voltage is generated.
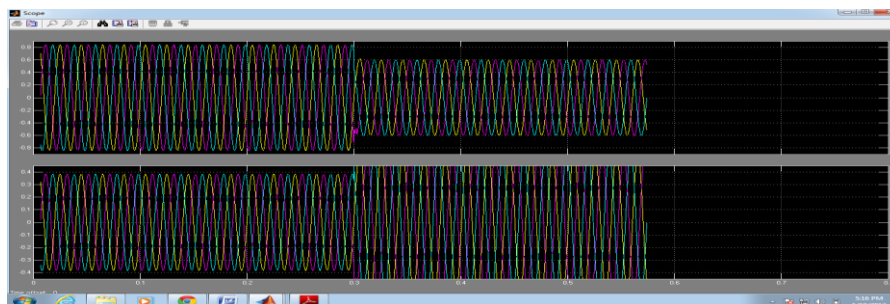
The PI controller thus gives the angle _. Control voltage at fundamental frequency (50 Hz) is generated and compared with the carrier frequency triangular wave of carrier frequency 1.5 kHz. The PWM pulses now drive the IGBT switch. The simulation modeling of PWM switched autotransformer used as mitigating device along with its control circuit is shown The autotransformer rating in each phase is 6.35/6.35 kV (as linevoltage is 11 kV) with 1:1 turns ratio. The effective voltage available at the primary of autotransformer is suchthat the load voltage is maintained at desired rms value (6.35 kV or 1 pu).
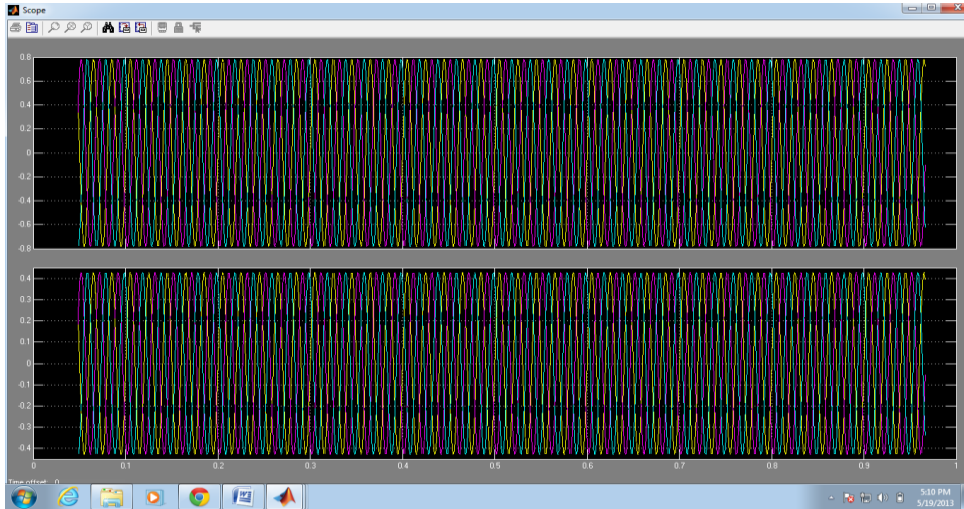
Voltage sag is created during the simulation by sudden application of heavy load of P=10MW and Q=50Mvar for a period of 0.1 sec(5cycles) from t1 = 0.1 sec to t2 = 0.2 sec. Shows the simulation waveforms of the load voltage for voltage sag of 27%.

X. CONCLUSION

A new voltage sag compensator based on PWM switched autotransformer has been presented in this paper. Control circuit based on rms voltage reference is discussed. The proposed technique could identify the disturbance and capable of mitigating the disturbance by maintaining the load voltage at desired magnitude within limits. The proposed technique is simple and only one IGBT switch per phase is required. Hence the system is more simple and economical compared to commonly used DVR or STATCOM. Simulation analysis is performed for 27% voltage sag for three phase system and simulation results verify that the proposed device is effective in compensating the voltage sag disturbances

REFERENCES

[1]   IEEE Recommended Practice for Emergency and Standby Power Systems for Industrial and Commercial Applications", vol. 445. The IEEE Standards Association, 1995.
[2]   R. Arnold. Solutions to the power quality problem. Power Engineering Journal, 2001, 65–73.
[3]   R. Cao, J. Zhao, et al. Series power quality compensator for voltage sags, swells, harmonics and unbalance. IEEE/PES Transmission and Distribution Conference and Exposition, 2001, 1(28): 543–547.
[4]   J. Chen, S. Song, Z. Wang. Analysis and implement of thyristor-based statcom. in: International Conference on Power System Technology, 2006, 1–5.
[5]   D. Lee, T. Habetler, et al. A voltage sag supporter utilizing a pwm-switched autotransformer. IEEE Transactions on Power Electronics, 2007, 22(2): 626–635.
[6]   C. Fitzer, M. Barnes, P. Green. Voltage sag detection technique for a dynamic voltage restorer. IEEE Transportation Industry Applications, 2004, 40(1): 203–212.
[7]   H. Masid, N. Moriun. Design of a prototype d-statcom for voltage sag mitigation. in: Proceedings of the 2004 National Power and Energy Conf, Kuallampur, Malaysia, 2004, 61–66.
[8]   M. McGranaghan, D. Mueller, M. Samotyj. Voltage sags in industrial systems. IEEE Transactions on Industry Applications, 1993, 29(2): 397–403.
[9]   G. Reed, M. Takeda, I. Iyoda. Improved power quality solutions using advanced solid-slate switching and static compensation technologies. in: IEEE Power Engineering Society 1999 Winter Meeting, vol. 2, New York, USA, 1999, 1132–1137.
[10]   B. Singh, A. Adhya, et al. Analysis and implement of thyristor-based statcom for small isolated alternator feeding distribution system. International Conference on Power Electronics and Drive Systems, 2006, 1: 274–279.
[11]   G. Taylor. Power quality hardware solutions for distribution systems: Custom power. in: IEE North Eastern Centre Power Section Symposium, Durham, UK, 1995, 11/1–11/9.

Introspection, Updates and Belief Revision as Agent Processes

Fernando Zacarias Flores[1,] Rosalba Cuapa Canto[2] Erick Madrid[3]

1,3 Computer Science – Autonomous University of Puebla,
2 Architecture Faculty – Autonomous University of Puebla

ABSTRACT:

A novel way of building intelligent agents is deployed. This proposal considers two aspects: In the first instance, our agents have a novel mechanism for updates, simple and easy of calculating and that guarantee our agents to stay always consistent. This mechanism was introduced and formalized in [13] and it is supported by the paradigm Answer Set Programming (ASP). ASP programs are written in the language of AnsProlog and its extensions [5, 11, 12, 14]. And, in a second plane, we generalize our kind of programs accepted in [13] and we accept general clauses, disjunctive clauses and augmented clauses. Considering these two aspects, we can develop systems to approach in a direct way to intelligent behavior.

KEYWORDS: *Agents, ASP, Belief revision, Introspection, Updates, AnsProlog, Consistent.*

I.   INTRODUCTION

The agent paradigm has recently increased its influence in the research and development of computational logic-based systems. A clear and correct specification is made through Logic Programming (LP) and Non-nomotonic reasoning that has been brought (back) to the spot-light. Also, the recent significant improvements in the efficiency of ***LP*** implementations for Non-monotonic Reasoning [3] have helped to this resurgence. However, when we develop a real application, we need a friendly front-end for user. For this reason, we integrated both Answer Set Programming (ASP - defined in [5]) and Java (object-oriented programming language), eliminating the traditional high gap between theory and practice. We make use of several important results [12, 13, and 15] that we have obtained in the last years in the field of non-monotonic extensions to Answer Set Programming. This can represent an important added value to the design of rational agents. Let us remember that ASP is the realization of much theoretical work on Non-monotonic Reasoning and Artificial Intelligence (AI) applications of *LP* for the last 15 years. The two most known systems that compute answer sets are DLV[1]and SMODELS[2].As we mention previously, our proposal is based in integrating to our agents a novel mechanism (given in section 3, definition 2) for updates. We want our agent to stay consistent in all moment, so that it acts in a correct and opportune way. We introduced and formalize this mechanism in [14], it is supported by ASP and its extensions [5, 12, 13, and 15]. In this context in our proposal we make use of two principles pointed out by Daniel Kahneman, *Nobel economy reward 2002*. First, people expected samples to be highly similar to their parent population and also to represent the randomness of the sampling process [16, 17], and second, people often rely on representativeness as a heuristic for judgment and prediction [8, 9].

Kahneman is a pioneer on the integration of economy and psychology research about making decisions. His job has opened a new line of studies, discovering how the human judgment sometimes takes short cuts and amazing paths which are very different from probability basic principles and theories about complex reasoning. All this opens a new line of studies on the subject of reasoning in logic, where most of the times we want to develop reasoning and complex mechanisms. However, Kahneman's studies show the opposite. The taking of decisions escapes many times from probabilities, economy predictions and from reasoning. For this reason, in our proposal we incorporate to our rational agents an update process with a human behavior through our definition introduced in [14]. In the same way as in [14], given a theory T, its knowledge is understood as the formulas F such that F is derived in T using intuitionistic logic. This makes sense, since in intuitionistic logic according to Brouwer, *A* is identified with "I know *A*". We consider that an agent whose knowledge base is a theory T believes F if and only if F belongs to every intuitionistically complete and consistent extension of T by adding only negated literals. Take for instance: ¬a → b. The agent knows ¬a → b, ¬b → ¬¬a and so on. The agent does not know however "a". Nevertheless, one believes more than one knows.

But a cautious agent must have its beliefs consistent to its knowledge. This agent will then assume negated literals to be able to infer more information. Thus, in our example, our agent will believe ¬a, and so it can conclude b. It also makes sense that a cautious agent will believe ¬a or ¬¬a rather than to believe "a" (recall that a is not equivalent to ¬¬a in intuitionistic logic). This view seems to agree with a point of view stated by Kowalski, namely ``that Logic and Logic Programming need to be put into place: Logic within the thinking component of the observation-thought-action cycle of a single agent, and Logic Programming within the belief component of thought'' [7]. See [12, 14, and 15] for a complete discussion on the role of intuitionistic logic in our approach, our proposal is focused to a real application.

On the other hand, our update process is safe and it maintains their knowledge base consistent, all this with the purpose that our agent gives reliable answer and in the right time. Later on, in a transparent way for the user, our agent carries out an introspection process. This process allows to the agent to refine their knowledge base, eliminating possible redundancies and restoring those independent beliefs to the new acquired knowledge. We use this introspect process supported by Kahneman's ideas. However, the question now arises is whether the result of an update process will depend on the particular set of sentences in the knowledge base, or only on the worlds described by this. We are interested in proposals that satisfy the fundamental principle of Irrelevance proposed by Dalal of syntax, that is, the meaning of the knowledge that results from an update must be independent of the syntax of the original knowledge, as well as independent of the syntax of the update itself. However, in our implementation we propose to reconsider the AGM postulates [1] under our new interpretation that considers "knowledge" and "belief". We use a new postulate which we call "Weak Irrelevance of Syntax" (WIS) and that we defined in [14]. This postulate, suggested by several authors [1, 4, 6], is satisfied by our update operator, as desired.The remainder of the paper is structured as follows: In section 2 we briefly recap the basic background used throughout the paper. In section 3, we present our new proposal on modeling of agents. Next, we present our application based in rational agents presented in section 4. Finally, in section 5, we give our conclusions and future work.

II.   BACKGROUND
In this section, we give some general definitions for our theory. We define our theory about logic programs.

2.1. Preliminary
Rules are built from propositional atoms and the 0-place connectives T and ⊥ using negation as failure (¬) and conjunction (,). A rule is an expression of the form:
$$\text{Head} \leftarrow \text{Body}$$
(1)
If Body is T then we identify rule (1) with rule Head. If a Head is ⊥ then we identify rule (1) with a restriction. A program is a set of rules. A logic program P is a (possibly infinite) set of rules. For a program P, I is a model of P, denoted $I \models P$, if $I \models L$ for all $L \in P$. As it is shown in [2], the Gelfond-Lifschitz transformation for a program P and a model $N \subseteq B_P$ (B_P denotes a set of atoms that appear in P) is defined by
$$P^N = \{rule^N : rule \in P\} \quad \text{where} \quad (A \leftarrow B_1, \dots, B_m, \neg C_1, \dots, \neg C_n)^N \text{ is either:}$$
a) $A \leftarrow B_1, \dots, B_m$, if $\forall j \leq n: C_j \notin N$;
b)   T,   otherwise
Note that P^N is always a definite program. We can therefore compute its least Herbrand model (denoted as M_{PN}) and check whether it coincides with the model N which we started with:

Definition 1. [2] N is a stable model of P iff N is the minimal model of P^N

To avoid technical problems we omit a formal definition of ASP for arbitrary formulas with two kinds of negations and all the theoretical concepts of answer sets see [12, 13, and 15].

2.2. Extending our kind of programs
We use several kinds of clauses found in literature [12]. A free clause is built from a disjunction of literals in the head and a conjunction of literals in the body. Such that a clause has the form h_1 v … v $h_n \leftarrow b_1$, … , b_m. where each h_i and b_j is a literal. Either the head or the body of a free clause could be empty to denote a constraint or a fact. A general clause is a free clause that does not allow negation in the head, all literals in the head of the clause should be positive atoms. Finally, a disjunctive clause is a general clause with a non-empty head, i.e. it is not a constraint. Next, we give an example using a disjunctive program in the context of our application; it consists on a research environment for our scientific community.

Example 1. In our system is common to establish an appointment, let us see the following example:
← appointment(X), schedule-available(X),spot(X).
spot(a).
spot(b).
~schedule-available(X) ← appointment(X), spot(X).
schedule-available(a) v schedule-available(b).

The interpretation is as follows: the first rule says that it is not possible to have an appointment, a spot and a schedule-available simultaneously. The third rule says that you don't have scheduled available X if you have an appointment X and a spot X. The last rule represents our schedule available, in this case, schedule-available(a) or schedule-available(b). As we can see, applying our definition given in the section 3 to this example we obtain that the program has two models: {schedule-available(a), spot(a), spot(b)} or {schedule-available(b), spot(a), spot(b)}, as desired. If our agent receives the following information: appointment(a) then, applying our definition 2 we obtain one model {schedule-available(b), spot(a), spot(b), appointment(a)}.

If later on we update with appointment(b), applying our definition again we obtain:
← appointment(X), schedule-available(X), spot(X).
spot(a)← ¬~spot(a). spot(b)←¬~spot(b).
~schedule-available(X) ← appointment(X), spot(X), ¬schedule-available(X).
schedule-available(a) v schedule-available(b)←¬~schedule-available(b).
schedule-available(a) v schedule-available(b) ¬~schedule-available(a).
appointment(a) ← ¬~appointment(a).
appointment(b).
Whose answer set is: {appointment(a), appointment(b), spot(a), spot(b), ~schedule-available(a), ~schedule-available(b)} as desired. This means that the rule five disappear, i.e., this rule doesn't have effect.

III. MODELING OF INTELLIGENT AGENTS

In this section, we present how our agents act in a dynamic environment. Before such situation, our agents should act in a correct and sure way, giving answers in an opportune way. This process is known as update process. We want this process to allow our agents to stay consistent in all moment. This guarantees us that our agents can always act in a reliable way.

3.1. Update definition

As part of our agents, we give our definition about update process. This definition was introduced in [14] and satisfies several properties of AGM postulates [1]. This gives to our agents an added value with respect to other proposals that don't satisfy them. One of the main aims of logic-constrained revision is to characterize suitable update operators through postulates like those formulated by AGM. In [4], the authors recapture these postulates and give their interpretation about AGM postulates in the update context. However, no such set of postulates would be adequate for every application [6]. Next, we present our update definition introduced in [14] and at once we present our new form of modeling agents. This form includes a novel mechanism that consists of the following three processes: Expansion, Update and Introspection.

Definition 2. Giving an update of two programs $P_\otimes = (P_1, P_2)$ over a set of atoms A, we define the update program $P_\otimes = P_1 \otimes P_2$ over A* consisting of the following items:
   (i)   all constraints in $P_1 \cup P_2$;
   (ii)   for each r ∈P_1;   L ← B(r), ¬~L. if H(r) = L;
   (iii) all rule r ∈P_2.

As we can see, our proposal is inspired by both AGM postulates and the proposal presented in [4]. As it is shown in [4] the interpretation in belief revision and in update coincides in some of the postulates. Also, our proposal (definition 2) coincides with [4] in a wide family of programs. It is necessary to highlight the simplicity of our proposal, that allows our agent to be able to respond in a correct and opportune way, for later apply our introspection process. Our agent faces the problem of update it knowledge base with new information considering three aspects: First, if new knowledge of the world is somehow obtained, and it doesn't have conflicts with previous knowledge then this only expands knowledge (we will refer it as expansion [1]). Second, if, on the contrary, new knowledge is inconsistent with the previous knowledge, and we want knowledge to be always consistent so that our agents can act in all moment, we should solve this problem somehow. We point out that new information is incorporated into the current knowledge base subject to a causal rejection principle,

which enforces that, in case of conflicts among rules, more recent rules are preferred and older rules are overridden. Third, once the agent applies our definition 2 formalized in [14] it can respond in a quick and opportune way, and our agent is in a new state. Since this moment our agent can apply its introspective process. This introspective process will allow our agent to revise the implications that our update could have generated. For instance, recapturing the example 1 we can see that is not necessary to weaken all rules, for example: spot(*a*) and spot(*b*).

3.2. Expansion
Next, we present a first example that shows the expansion process of new knowledge, i.e., if new knowledge of the world is somehow obtained, and it doesn't have conflicts with the previous knowledge then this only expands knowledge base.

Example 2. The next example shows the expansion process that our agent executes when to he/she is presented knowledge without consistency problems with previous knowledge base.

Let P be: r1: mail-revision ← computer-on, internet-on.

                    r2: another-task ← ¬mail-revision.

                    r3: computer-on.

                    r4: internet-on ← computer-on, ¬power-failure.

Let P_1 be: r5: class-preparation ← ¬there-are-students.

                    r6: research ← ¬busy.

In this case, as we can see the new knowledge doesn't have conflicts with previous knowledge. In this case, our agent only applies the expansion process. Then, the answer set of $P \otimes P_1$ is: {computer-on, mail-revision, internet-on, class-preparation, research} as desired.

3.3. Update
Next, we recapture the previous example and show our update process using our definition 2, which coincides in a wide family of programs with the one presented in [4].

Example 3. This example shows the update process. This happens if by the contrary, new knowledge is inconsistent with the previous knowledge base, and we want that our agent maintains its knowledge always consistent so that our agents can act in all moment.

Let P_1' as follows: r7: ~computer-on ← power-failure.

                        r8: power-failure.

Then, applying our previous definition 2 we obtain the following program:

                    r1: mail-revision ← computer-on, internet-on, ¬~mail-revision.

                    r2: another-task ← ¬ mail-revision, ¬~another-task.

                    r3: computer-on ← ¬~computer-on.

                    r4: internet-on ← computer-on, ¬power-failure, ¬~internet-on.

                    r7: ~computer-on ← power-failure.

                    r8: power-failure.

The answer set of $P \otimes P_1$' is: {~computer-on, power-failure, another-class} as desired.

As we can observe, in this example the rules r7 and r8 have problems of consistency with previous knowledge. In this sense, the most recent rules are preferred and oldest ones are overridden. This proposal gives to our agents a quick and efficient mechanism.

3.4. Update
Here, we present our proposal related with update process called "introspective process", that it can be seen more near to belief revision. This process is executed for our agent after having applied our update process. Also, this process can be used to analyze short messages in mobile telephony. Next, we recapture the previous example and consider that our knowledge base P includes the rules from r1 to r6, and the new knowledge consists of rules r7 and r8. The first step consists in applying our update process. However, the rules r5 and r6 are independent of the rest of the program. Our agent will make an introspective analysis evaluating if it was necessary to weaken all the rules, made in the update process. This is an excellent process that our agent has and that allows itself to restore all those rules that are not involved with the new knowledge (in this case, the rules r5 and r6).

Example 4. This example shows the introspective process. This happens after applying our definition 2. Then, the answer set of P ⊗ P$_1$' is: {~computer-on, power-failure, another-class, class-preparation, research} as desired.

Our conceptual solution to modeling introspection is as follows:

- Keep a history of the last *n* updates done by our agent as well as its original theory representing its sets of beliefs.
- Verify step by step each update. Here, we have two major concerns. First, to correct only the ``relevant'' rules (recall in example 3 that we modified all rules). Second, if the belief base becomes inconsistent (a non-typical situation) we need to apply a form of belief revision in order to achieve consistency. We have partial theoretical results reported in [11, 12, 13, 14, and 18].
- If a major problem can not be resolved by (2). Our idea is that the system should request a discussion group (by e-mail) with others agents (possibly humans) such that all together can find a solution.

However, is required to continue working in this research line.

IV. Application based on intelligent agents

In this section we present our application based in rational agents. This proposal integrates both ASP and Java, eliminating the traditional high gap between formal theory and practice. We use java as front-end, in which we develop the interfaz for user and the database administration. While in the declarative part, we use answer set programming through DLV. This paradigm gives the formal support to our agents. Nowadays, the agent paradigm has recently increased its influence in the research and development of computational logic based systems. However, the development of "rational" agents is a hard task, due to this involves both the updates and beliefs revision process. Besides, when we think in rational agents, we should consider some human characteristics like: reasoning, planning, and acting in a dynamic world. Our application consists in a work environment (figure 1) that integrates necessary tools for correct performance of our scientists, incorporating a new formalization about update process. Our application consists of: an intelligent calendar that has as main ingredient the negotiation of meetings among multiple members of our scientific community via rational agents; a chat, that allows virtual meetings reducing the big distances; electronic mail and short messages via mobile telephone, used as our general communication channel in the negotiation meetings by our agents; an intelligent editor, that counts with a rational agent that assists in papers writing and a knowledge base on the scientists' belief. In the figure 1, we present the main interfaz of our application. As we can observe one of the main tasks is the presentation of the schedule of each one of the researchers in our community. All this with the following purpose: that any member of the community can request a meeting with any of the researchers. The acceptance or not of a meeting depends on the beliefs that our agent has with respect to the required researcher. These knowledge bases are formed for beliefs, knowledge and preferences. Obviously, these knowledge bases change in a dynamic way according to the researcher's dynamic behavior.

Example 5. This example shows how our agent acts when it faces new knowledge. Suppose that knowledge base is as follows:

workday(mon).   workday(tue).   workday(wed).   workday(thu).   workday(fri).

coffee-hour(1600,Day) ← workday(Day).

call-home(1930,Day) ← workday(Day). research-time(1200,1400,mon,alone).

research-time(1200,1400,wed,group). research-time(1600,1800,fri,alone).

superior(Super, Person) ← boss(Super, Person).

superior(Super, Person) ← boss(Super, Someone), superior(Someone, Person).
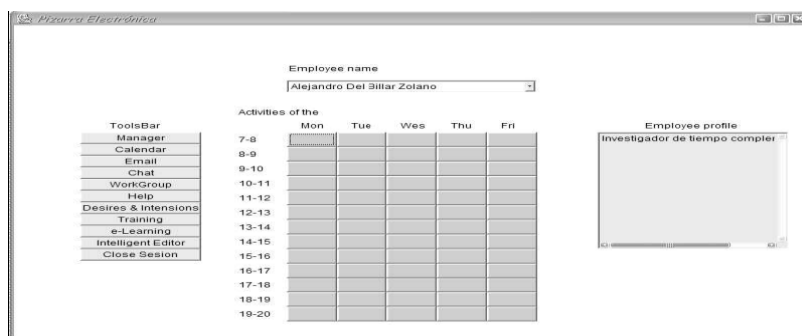
Figure 1. Electronic board

Now, suppose that our agent receives the new following information:

~coffee-hour(1600,Day) ← workday(Day).   tee-hour(1600,Day) ← workday(Day).

In this case, our agent doesn't apply the expansion process, because the new information has conflicts with the previous information. Therefore, applying our definition about belief revision we obtain:

workday(mon) ← ¬~workday(mon).   workday(tue) ← ¬~workday(tue).
workday(wed) ← ¬~workday(wed).   workday(thu) ← ¬~workday(thu).
workday(fri) ← ¬~workday(fri).
coffee-hour(1600,Day) ← workday(Day), ¬~coffee-hour(1600,Day).
call-home(1930,Day) ← workday(Day), not ~call-home(1930,Day).
research-time(1200,1400,mon,alone) ← ¬~research-time(1200,1400,mon,alone).
research-time(1200,1400,wed,group) ← ¬~research-time(1200,1400,wed,group).
research-time(1600,1800,fri,alone) ← ¬~research-time(1600,1800,fri,alone).
superior(Super, Person) ← boss(Super, Person), ¬~superior(Super, Person).
superior(Super, Person) ← boss(Super, Someone), superior(Someone, Person),
¬~superior(Super, Person).   ~coffee-hour(1600,Day) ← workday(Day).
tee-hour(1600,Day) ← workday(Day).

In this case, the answer set is:
{workday(mon), workday(tue), workday(wed), workday(thu), workday(fri), ~coffee-hour(1600,mon), ~coffee-hour(1600,tue), ~coffee-hour(1600,wed), ~coffee-hour(1600,thu), ~coffee-hour(1600,fri), call-home(1930,mon), call-home(1930,tue), call-home(1930,wed), call-home(1930,thu), call-home(1930,fri), research-time(1200,1400,mon,alone), research-time(1200,1400,wed,group), research-time(1600,1800,fri,alone), tee-hour(1600,mon), tee-hour(1600,tue), tee-hour(1600,wed), tee-hour(1600,thu), tee-hour(1600,fri)} as desired.

As we can observe, the new information is related with the coffee hour only. Then, our agent can apply the introspection process, i.e., our agent can execute its introspective process. Considering that it is totally unaware to the new knowledge. In this case, we can consider following knowledge obtaining:
research-time(1200,1400,mon,alone).   research-time(1200,1400,wed,group).
research-time(1600,1800,fri,alone).   superior(Super, Person) ← boss(Super, Person).
superior(Super, Person) ← boss(Super, Someone), superior(Someone, Person).
On the other hand, our agent should make a more exhaustive analysis with respect to the most relevant knowledge with the purpose of not modifying the original semantics.

V.   CONCLUSIONS AND FUTURE WORK
In this paper, we considered a new proposal for modeling agents. In this proposal we include three very important characteristics that give to our agent a human behavior. This behavior is very similar to the human behavior considering Kahneman's ideas. Humans don't always make very exhaustive reasoning, mainly in situations where we should give answers in a quick, opportune, and correct way. Also, our application is based in rational agents whom lean in an update process that is independent from syntax and whose theoretical support is presented in [14]. With respect to future work, we will continue our research in belief revision. In particular form, we will continue our analyses about introspective process. Because, we consider that this process is adapted for when one wants that our agents to give opportune answers in time and it forms. Also, another future objective is to endow our agents of new capacities that allow them to analyze the short messages among mobile telephones, with the objective of determining a reception of 100 percent of them.

REFERENCES
[1]   C.E. Alchourron, P. Gardenfors, and D. Makinson. On the logic of Theory Change, Partial Meet Functions for Contraction and Revision Functions. Journal of Symbolic Logic, 50:510-530, 1985.
[2]   G. Brewka, J. Dix, and K. Knonolige, nonmonotonic reasoning, an overview. CSLI Publication Eds. Leland Stanford Junior University, 1997.
[3]   D. De Schreye, M. Hermenegildo and L.M. Pereira. Paving the Roadmaps: Enabling and Integration Technologics. http://www.compulog.org/net/Forum/Supportdocs.html.
[4]   T. Eiter, M. Fink, G. Sabattini, and H. Thompits. Considerations on Updates of Logic Programs. In M.O. Aciego, LP. de Guzmßn, G. Brewka, and L.M. Pereira, editors, Proc. Seventh European Workshop on Logic in Artificial Intelligence JELIA 2000, vol. 1919 in Lecture Notes in Artificial Intelligence. LNAI, Springer 2000.
[5]   M. Gelfond, V. Lifschitz. The stable model semantics for logic programs. Proceedings of the Fifth International Conference on Logic Programming 2, MIT Press. Cambridge, Ma. pp.1070-1080.
[6]   H. Katsumo and A.O. Mendelzon. On the difference between updating a knowledge base and revising it. in: J.A. Allen, R. Fikes and E. Sandewell. eds.. Principles of knowledge representation and reasoning: Proceedings of the Second International Conference (Morgan Kautmann. San Mateo. CA. 1991) pp. 387--394.

[7]   R. Kowalski. Is logic really dead or just sleeping. In Proceedings of the 17th International Conference on Logic Programming, pages 2-3, 2001.

[8]   D. Kahneman and A. Tversky. Subjective probability: A judgment of representativeness, Cognitive Psychology, vol 3, pp. 430--454, 1972.

[9]   D. Kahneman and A. Tversky. On the Psychology of prediction, Psychological Review, vol. 80, pp. 237-251, 1973.

[10]   V. Lifschitz, D. Pearce, and A. Valverde. Strongly equivalent logic programs. ACM Transactions on Computational Logic, 2:526-541, 2001.

[11]   M. Osorio and J.C. Acosta. Towards modelling an intelligent calendar agent with LUPS, IASTED International Conference on Artificial Intelligence and Applications, (AIA'2003) Innsbruck, Austria, February, pp. 10--13, 2003.

[12]   M. Osorio, J.A. Navarro and J. Arrazola. Applications of Intuitionistic Logic in Answer Set Programming, Journal of Theory and Practice of Logic Programming, TPLP , 2003.

[13]   M. Osorio, J.A. Navarro, and J. Arrazola. Equivalence in Answer Set Programming, Proceedings of LOPSTR 01, LNCS 2372, pp. 57-75, Springer-Verlag, Paphos, Cyprus, November 2001.

[14]   M. Osorio and F. Zacarias, Irrelevance of Syntax in updating answer set programs, Proceedings Of Fourth Mexican International Conference On Computer Science Enc'03, pp.183-188, Eds. J. H. Sossa, and E. Perez, México, 2003.

[15]   D. Pearce. From Here to There: Stable negation in Logic Programming, in D. Gabbay, H. Wansing (Eds.) What is Negation? Kluwer Academic Publishers, Dordrecht,.

[16]   A. Tversky and D. Kahneman. Belief in the law of small numbers. Psychological bulletin, Vol. 2, pp. 105--110, 1971.

[17]   A. Tversky and D. Kahneman. Judgment Ander uncertainty: Heuristics and biases, American association for the Advancement of Science, vol. 1, pp. 1124--1131, 1971.

[18]   Fernando Zacarías, Mauricio Osorio, José Arrazola. Updates based on Structural Properties –USP-. Gests international transactions on computer science and engineering, pp. 61-72, issn: 1738-6438, isbn: 89-953729-5-8, October 2005.

.

Waypoint Navigation System Implementation via a Mobile Robot Using Global Positioning System (GPS) and Global System for Mobile Communications (GSM) Modems

Sulaiman Khan [1,] Kashif Ahmad[2], Mohsin Murad[3], Imran Khan[4]

[1,2,3] Department of Computer Systems Engineering, University of Engineering and Technology Peshawar, Pakistan,
[4] Department of Mechanical Engineering, University of Engineering and Technology Peshawar, Pakistan

ABSTRACT:
This paper presents the use of Global Positioning Systems (GPS) as geographic information and navigational system for a ground based mobile robot. The proposed mobile robot contains a GPS system for navigation and sensors (LV-MaxSonar–EZ4) for obstacle avoidance system, and GSM Modem for communicating with user. The Mobile robot navigates to the waypoint specified by the user through the GSM Modem and avoids the obstacles in its way to destination. The experiments are carried out in the university lawn with a test bed mobile robot for point-to-point motion using a GPS.

KEYWORDS: *Navigation System; GPS; GSM; Mobile Robot, Waypoints*

I.   INTRODUCTION

The navigation systems are concerned with the monitoring and controlling of the movements of a vehicle/craft in a physical space, and are very handy in situations where the human being can not operate or where it is hazard for the human beings to operate such as in mine detection and bomb disposal. Apart from the military usage, the artificial waypoint navigation systems are much effective in defining the artificial air ways. Based on its application scope, the navigation systems can be classified in different categories such as land, space, marine and aeronautic navigation systems [1]. The waypoints represent the sets of coordinates or the small regions that provide the salient information for the identification of a point in any physical space. The nature of co-ordinates is application dependent e.g. the longitude and latitude may be the possible co-ordinates set for a terrestrial navigation system. This set of co-ordinates provides the trajectory information to the robotic vehicle or craft to follow the correct route to destination.This work presents a low cost and efficient approach for chasing the co-ordinates given by the user to the Mobile Robot Vehicle using the GPS based navigation system. In order to make the system more accurate and visual the proposed work utilizes the Google Maps, which is a free web mapping service application and technology provide by Google. The main task of such mobile robots is to navigate from one co-ordinate to other co-ordinate avoiding the obstacles in its way and give feedback to the user.

II.   RELATED WORK

The navigation systems are concerned with the monitoring and controlling the movements of a vehicle/craft in a physical space. For tracking the route, the navigation system utilizes the waypoints information such as presented by Millington et al. [2], in which they utilized the waypoint information for the navigation system in a vehicle. The co-ordinates information are used as route, and the waypoints information and the vehicle position are displayed on the screen with a feedback system for controlling the vehicle moments directions. The use of Global Positioning System (GPS) for the creation of waypoints i.e. set of co-ordinates in physical space for the navigation systems have been proved very effective. In Global Positioning System (GPS) based navigation systems, the trajectory of the vehicle or craft is determined through a series of waypoints and followed by navigating the next waypoints until the destination is reached such as presented in [3,4]. hoi et al. [3] provided autonomous mobile robot using GPS, the robot follows its trajectory with feedback through GPS receiver, and evade obstacles with the help of photo-sensors. Along with this, one can trace a Mobile robot with a wireless RF communication module. Hamid et al. [4] present another navigation system's implementation in a mobile robot in a different way using GPS for navigation and sonar sensors for obstacle avoidance. The beauty of this work is that they are using command loop daisy changing application method.
The similar work is presented in by sukkarieh et al. [5] for land vehicle applications. Their work combines the inertial measurement unit with global positioning system for the enhancement of integrity of the navigation

system by considering both the low frequency faults and high frequency faults in inertial measurement unit and PGS respectively during and before the fusion of inertial measurement unit and GPS. Bruch et al. [6] represent a land based navigation system for vehicles implemented on Man Portable Robotic System Urban Robot. This navigation system uses combination of Kalman filter, waypoint and some inexpensive sensors along with the GPS and implemented on an embedded processor.

III. ROBOTIC PRINCIPLE

To maximize the accuracy for the target position of the mobile robot, such an algorithm is needed that uses the basic principles of GPS receiver. Information about the position in the form of a string is loaded from the GPS receiver to the microcontroller through serial port. As the Garmin eTrex-H GPS is used in this work, the output string format is given in Figure 5. The mobile robot is sending that information through GSM modem to the system where it is received serially. The co-ordinates received are plotted in Google Map and on return sending targeted (desired) co-ordinates from the system through a GSM modem to the mobile robot, the GSM modem fixed in the mobile robot, gets the co-ordinates serially and compare that co-ordinates with the sent co-ordinates and travel to the position, by the way if there is an obstacle in the way it will be avoided through Sonar sensors.

IV. NAVIGATION

The transition of an artificial entity from one coordinate (position) to another is termed as navigation. The GPS based navigation system, originally developed for military purposes is formed from 24 satellites and is the most widely used navigation system all over the world. Nowadays, instead of the militarily usage, the GPS is used in health, communication and different other utilities related to motorcars, airplanes, ships, drones and could be apply to all other moving amenities. The figure 1 shows example of GPS based navigation system.
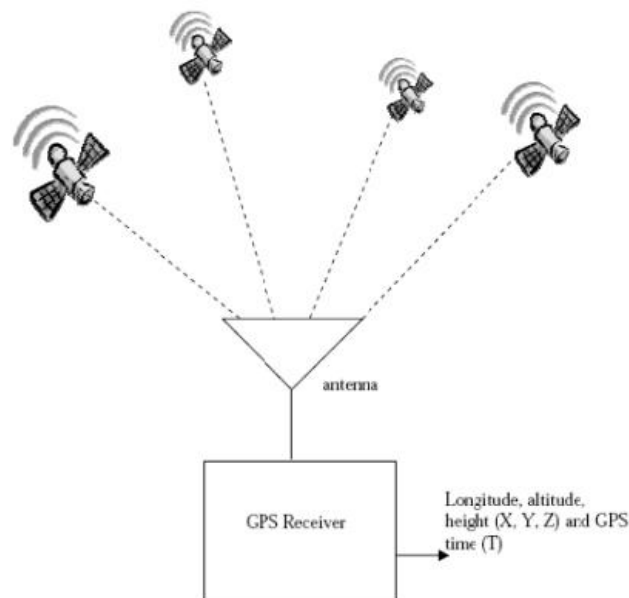
Fig.1. Determination of (x,y,z,t) from Four Coded Time Signal [7]

Many considerations have been made for the selection of GPS module for this particular application. After the study of all choices and considerations, the mobile robot designed for this application prefers the use of Garmin eTrex H for GPS receiver. The main reason is that the small size of the GPS receiver makes it light and this is 12 channel GPS receiver designed to operate with L1 frequency and the GPS receiver tracks and uses the signal to compute and update position. There are a number of output formats supported by this GPS receiver; however the proposed work uses the text output format. Text mode is a simple output mode that provides a string containing the information about time, position and velocity. An example is shown figure 2 given bellow.

```
@000607204655N6012249E01107556S015+00130E0021N001
8U0000
```
1st bit- show Start bit
2nd -12th bits show Time bits
13th -bit is Latitude Hemisphere ('N' or 'S')
14th -20th –bits shows Latitude position
21st –bit is Longitude Hemisphere ('E' or 'W')
22nd -29th –bits shows Longitude Position
30th –bit is Position Status
31st -33rd –bits shows Horizontal Position Error
34th –bit is Altitude Sign ('+' or '-')
35th – 39th –bits shows Altitude
40th –bit shows Velocity Direction ('E' or 'W')
41st -44th –bit shows Velocity magnitude(Meters per second in tenths)
45th –bit is Velocity Direction ('N' or 'S')
46th -49th –bit Velocity Magnitude(Meters per second in tenths)
50th –bit is Vertical Velocity Direction('up' or 'down')
51st -54th –bits shows Vertical velocity Magnitude(Meters per second in tenths)
55th -56th –bits shows Sentence End

Fig. 2. Text Output Formatof GARMIN eTrex H GPSReceiver

V.   OBSTACLE AVOIDANCE

During the course of navigation from initial to the final waypoint there may be an obstacle which must have to be avoided. To achieve this purpose this project takes the advantage of sonar sensors to avoid an obstacles in its way. This project focused on the use of LV-MaxSonar–EZ4. We are using 2 sonar sensors implemented in front of mobile robot and 2 sonar sensors on back each with 11cm apart from each other. The choosing of the LV- MaxSonar–EZ4 Sonar Sensors based on the specification of the sensors. LV-MaxSonar-EZ4 is a low power i.e. 2.5V– 5.5V supply with 2mA typical current draw. It is 42 kHz Ultrasonic sensor. The rate of reading is 20Hz. LV-MaxSonar -EZ4 detect object from 0-inches to 254-inches (6.45 meters) with 1-inch of resolution. The interface output formats included are pulse width output, analog voltage output, and serial digital output. LV-MaxSonar –EZ4 are very low cost sonar range finder with reliable and stable range data with a quality beam characteristics.
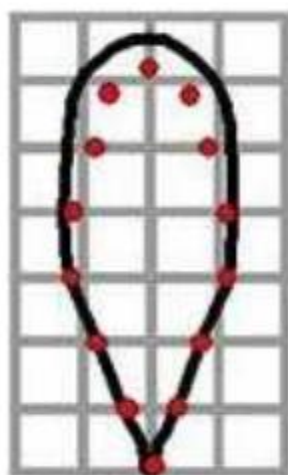
Fig. 3. Beam Charactristics Approximation of LV-MaxSonarEZ

As previously mentioned an extremely wide beam width is the characteristic of the most ultrasonic sensors. This is cause for concern, specifically when avoiding obstacles comes onto play. The beam width characteristic is shown in Figure 3. This shows the detection pattern for a 4-foot tall, 3.25-inch diameter post. Figure 3 is the demonstration of a problem resulting from wide beams, producing a large distance ambiguity making an accurate location of an obstacle near impossible. The Figure 4 shows four different scenarios a sensor with a wide beam ma face, each of which returning the same analog output creating obstacle location ambiguity. Pulse data can be converted to centimeter or inches form equation (1) and (2).

$$Distance\ (cm) = \frac{microsecounds}{29} \qquad (1)$$

$$Distance\ (inches) = \frac{microsecounds}{74} \qquad (2)$$

VI.   TARGET TECHNOLOGY

After detailed analysis of the proposed navigational system, the 89C51 is selected from the microcontroller suite for the implementation of the decision making module of the proposed mobile robot. 89C51 can execute 921,583 single-cycle instructions per second with a processing speed of 24 MHz. The on-chip Flash provides the reprogramming facility through both that the in-system and nonvolatile memory programmer. Instead of the high accuracy and flexibility, the 89C51 provides a cost effective solution in different applications.

The AT89C51 provides:

- ✓   4Kbytes Flash Memory
- ✓   128 bytes of RAM
- ✓   32 I/O lines
- ✓   Two 16-bit timer/counters
- ✓   Five vector two-level interrupt architecture
- ✓   On-chip oscillator
- ✓   Full duplex serial port
- ✓   Clock circuitry

Additionally it provides two software selectable power saving modes.
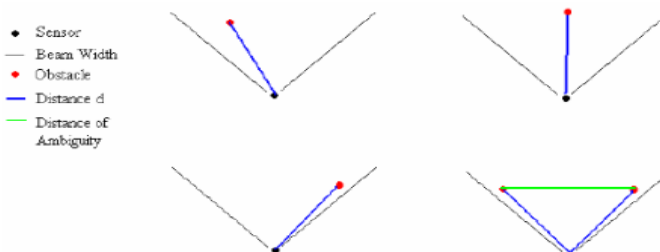
Fig.4. Different Scenarios returning the same Sensor Output [9]
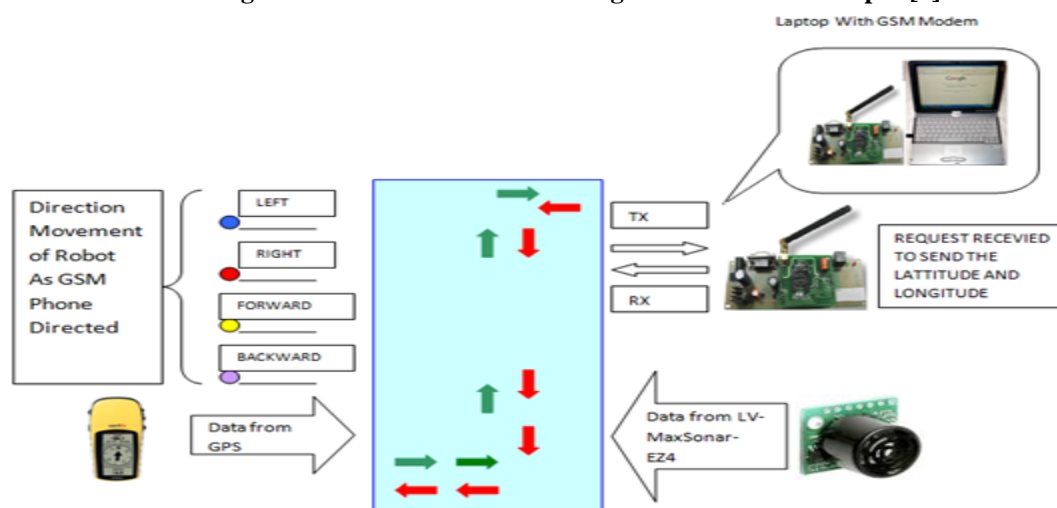
Fig.5. Block Diagram of Proposed Navigation System
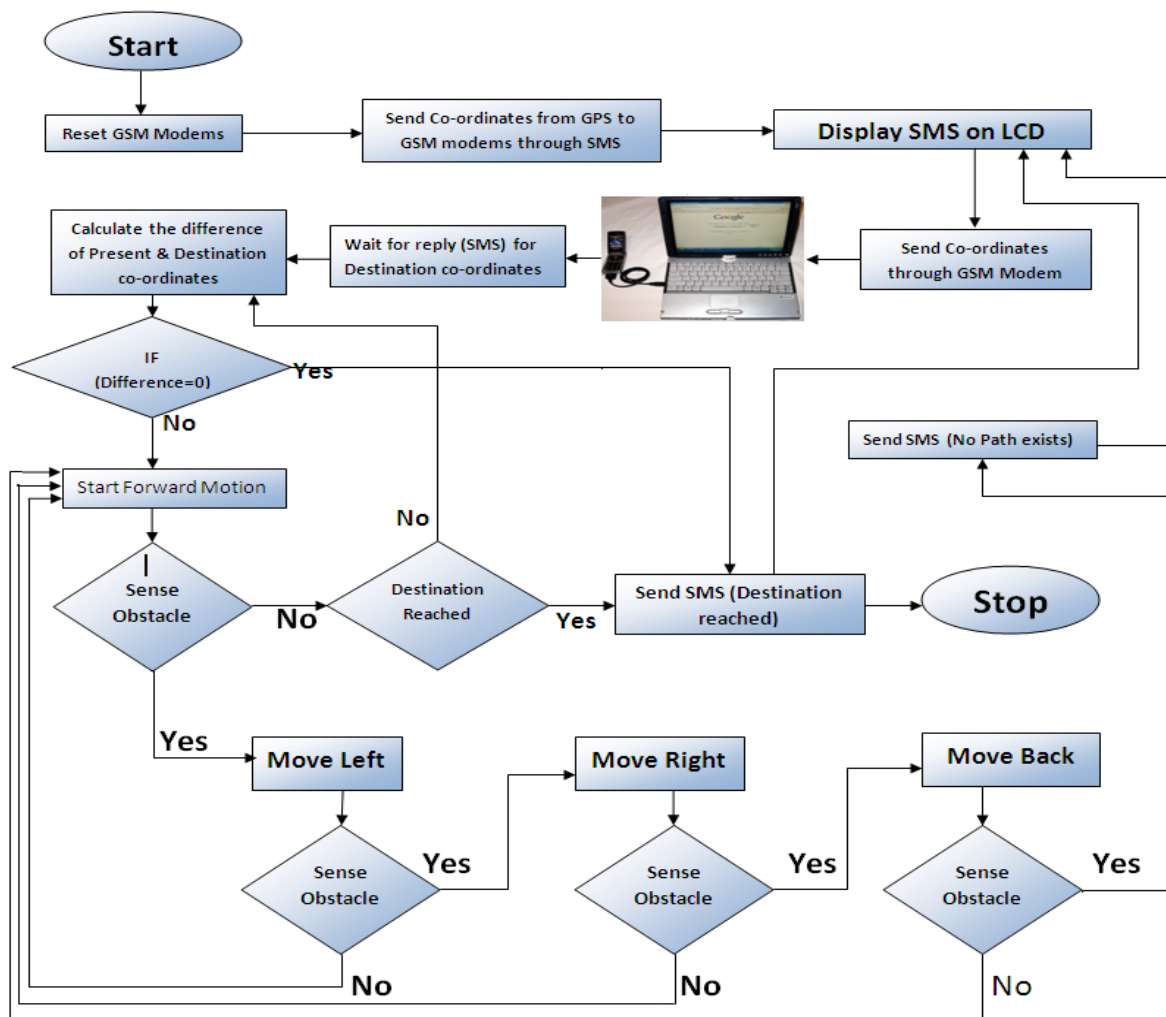
Fig.6. Flow Chart of the Proposed Decision Making Module

VII.   DECISION MAKING MODULE

As our Navigation Robot contains GPS device, GSM modems and main circuitry, so first we reset the whole circuitry along with GPS & GSM modems. After resetting the Robot sends its current GPS co-ordinates to the GSM modem connected with a Laptop. The destination co-ordinates can be calculated from the Google map or from Google earth software and send to the Robot through the GSM connected with the Laptop. When the Robot receives the destination co-ordinates it drags a line between its current location and destination point. Also after receiving the destination co-ordinates it calculate the difference between these co-ordinates and store the result, and do a comparison (if difference = 0) it means the destination reached and it stops, but (if difference ≠ 0) it continue its motion and moves towards its destiny.

Now during its motion if an obstacle appears in its way the it changes its direction, if obstacle comes in forward direction it sends the co-ordinate and moves to left and sense the obstacle, if there is no obstacle in left it moves towards its destiny but if obstacle exists then again it sends the co-ordinates and moves towards Right, again it sense the obstacle in its way, if there is no obstacle it drags a line between its current position and its destiny and start motion but if obstacle exists then it sends the co-ordinates to the GSM connected with Laptop and moves in backward direction, again it sense the obstacle if there is no obstacle it moves for one minute and then change its direction and moves towards it destiny, but if obstacle exists its sends co-ordinates along with message "No path exists" . The distinguishing and more interesting point in our Navigation Robot from other Navigation Robot is that, that when it sends a co-ordinate to the GSM modem connected with Laptop we see the track of the Navigation Robot as we represented that by an arrow. Due to this track we can easily deduce that an obstacle appeared in its way and it changed its direction. And another interesting point is that, that whenever it receives a co-ordinate from the GPS it calculates the difference and compares that result with previous result to check "whether the current difference is larger than the previous one it means the Navigation system moving far away from its destination point and if smaller it means it moves towards its destiny".

VIII.   DESIGNED MODEL

Figure 5 shows the block diagram of the proposed navigation system while figure 7 shows the structure of the mobile robot truck model. The robot model used in this work is in dimension of 48 cm in length, 24 cm in width and 16 cm in height. The mobile robot truck consists of GPS receiver, GSM modem, the Ultrasonic sensors in front and the circuitry. The mobile front wheel acts as a steering and the back wheels for force, so the motors for forcing and steering are inside the mobile robot. The biggest challenge was that how to supply power to each part of the robot. The ultrasonic sensors need 2.5 – 5.5Volt and the other IC's need 5 volt i.e. NE555N, Relay, MAX232ACPE and L293D. This model uses 7805 as a Voltage regulator which provides a convenient power source for most TTL components. Output signal from GPS receiver and GSM Modem is RS-232 signal that can be changed to TTL signal by using converting circuit as shown in Figure 3. L293d is a dual H-Bridge motor driver, so with one IC interface two DC motors can be interfaced. These motors can be controlled in both clockwise and counter clockwise-direction.
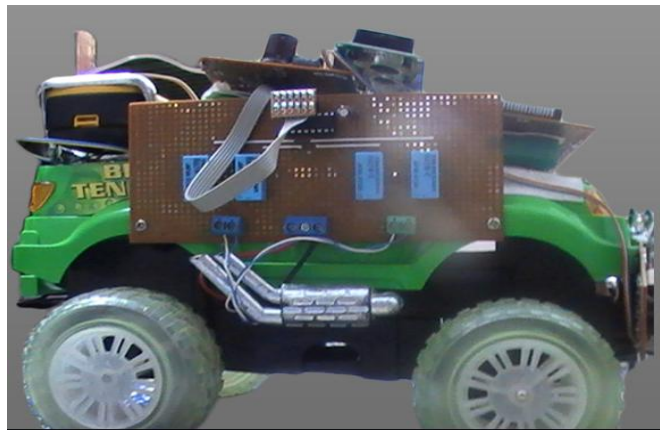
Fig.7. Model of the Proposed Robot

IX.   CONCLUSION

The convenience and high performance makes the Global Positioning System a better choice for the navigation of a mobile robot. The results of Global Positioning System (GPS) on the mobile robot are probable. Based on the performance results of the GPS in different scenarios during this work, it can be concluded that Compare to other methods of navigation, GPS is simple, highly accurate and weather proof.

REFERENCES

[1]   Rell Pros-Wellenhof, Bernhard (2007), "Navigation: Principles of Positioning and Guidance", Springer. pp. 5–6. ISBN 9783211008287.
[2]   Jeffrey Alan Millington, Chandiran Palanisamy, Rhonda Marie Paprocki et al. (2004), "Vehicle Nevagation System with Off-Road Navigation", United States Patent, Patent No. US 6,836,725 B2, date of Patent: Dec. 28, 2004.
[3]   Hawan-Seok Choi, Han-Sil Kim (2005), "Autonomous Mobile Robot Using GPS", proceeding on international conference of control and automation (ICCA2005) Budapest, Hungary, Vol. 2, Pp. 858-862.
[4]   MHA Hamid, AH Adom, NA Rahim, MHF Rahiman (2009), "Navigation of Mobile Robot Using GPS and Obstacle Avoidance System with commanded Loop Daisy Chaining Application Method", Proceeding on 5th International Colloquium on Signal Processing & its Application (CSPA) Kaula Lumpur, Vol. pp. 176-181.
[5]   Sukkarieh, S, Nebot, E.M., Durrant-Whyte, H.F (1999), "A high integrity IMU/GPS navigation loop for autonomous land vehicle applications", Robotics and Automation, IEEE Transactions on, Volume: 15, Issue: 3, pp-572-578.
[6]   Micheal H. Bruch, G.A. Gilbreath, J.W. Muelhauser, J.Q. Lum (2002), "Accurate Waypoint Navigation using non-differential GPS", AUVSI Unmanned-Systems.
[7]   H.R. Everett (1995), "Sensors for Mobile Robots", Naval Command, Control and Ocean Surveillance Center, A K Peters Ltd.
[8]   Paul Nuss, (2007), "Object Detection with Sensors".

.

Light Weight, Low Cost, Wearable Ecg Monitoring

Shraddha Parag Deopujari[1], Dr. Ashok Gaikwad[2]

[1,]Student, Department of Instrumentation and Control (Biomedical)[1]
[2,]Professor, H.O.D. Department of Instrumentation and Control[2]
Cummins College of Engineering, Pune, India

ABSTRACT

This paper describes the proposed signal processing scheme for wearable monitoring Electrocardiogram. This includes the acquisition, amplification and filtering of the on body ECG signal. EC signal is very sensitive in nature and if a small noise get mixed with the original signal, various characteristi of the signal changes. Hence filtering is an important issue. Remote monitoring of ECG and oth Physiological signals is becoming important as it can significantly reduce the cost and risks involved personal healthcare. Digital signal processing and data analysis are very commonly used methods, biomedical engineering research. Microcontroller based ECG monitoring is also described in this pape PIC16F877A microcontroller with inbuilt ADC is used for the purpose of R-peak detection and Heart ra monitoring. Wearable ECG monitoring is the need of ageing society now a days. Our device is low co monitor and easily wearable ECG monitor, so that it can help in home monitoring of ECG data. The basic a behind this project is detection of abnormality or emergency condition, before they prove dangerous.

KEYWORDS: *Electrocardiogram (ECG), Electrodes, Filters, Graphical User Interface, Heart ra monitoring, Noise, Signal Processing*

I. INTRODUCTION

World-wide survey conducted by World Health Organization (WHO) has confirmed that Heart related diseases are on the rise today. Many of these cardiac problems are due to modern lifestyle, food habits, obesity, smoking, lack of physical exercise, etc. Post operative patients can develop complication once they are discharged from the hospital. In some patients[1], cardiac problem may re-occur when they start doing their routine work. Hence ECG of such patients needs to be monitored continuously for some time after operation. The main challenge in the development of wearable device is local ECG signal processing with high accuracy. It is important to acquire the accurate ECG signal and remove the unwanted noise using a proper signal processing scheme with low expenditure. Various experiments are in progress to develop a good signal processing scheme and get a low cost wearable ECG monitoring device. These type of wearable monitoring devices must be run using the small batteries as the power supply for less consumption of power[2].

This reduces the weight and dimensions of the device. Also these must be easy to use unchanging the quality of life of the personal. A high level of integration is required to minimize the size and cost of such a sensor. It is desirable to do most of the signal processing tasks like ECG filtering and QRS detection locally. Once the QRS complex position is found, the other components of ECG signal are not difficult to be detected, such like P and T waves, ST segment. So the QRS complex detection is the key for ECG analysis[3]. This paper presents an ECG signal conditioning circuit. ECG signal is usually influenced by baseline drift, motion artifact, and power line interference.
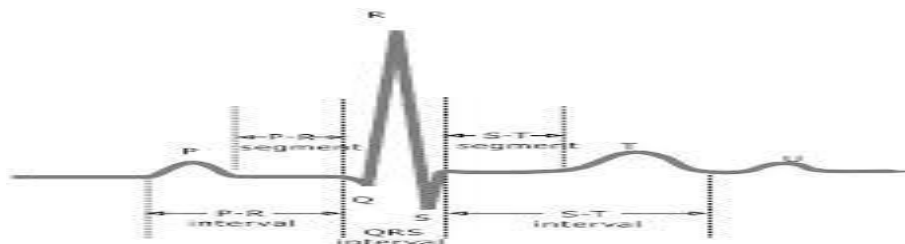
Figure1. Normal ECG wave with noted features

An ECG is a graphical display produced by an electrocardiograph, which records the electrical activity of the heart over time. The electrocardiograph as shown in Figure 1 is obtained by measuring the electrical potential between various points of the body. The ECG waveform consists of several sections, labeled P, Q, R, S, T and U. Each section has its own characteristics and yield important diagnostic information, which an experienced cardiologist can interpret to diagnose a wide range of possible ailments.

The paper is organized as follows. In section II gives the background behind the wearable monitoring. Details of electrodes used are given in section III. Section IV gives the Hardware design of the proposed scheme and simulation results are presented in section V. Concluding remarks are given in section VI.

II. BACKGROUND

From literature study it is found that, Electrocardiograph devices are the most widely employed tool used for cardiac monitoring. Two different classes of ECG devices are in general use today. The "standard" ECG generally involves the connection of between twelve and fifteen conductive leads to a patient's chest and/or extremities via adhesive pads[1]. The device records a short sampling of the heart's electrical activity between various pairs of electrodes. Each pair of leads provides a unique depiction of the heart's electrical impulses as they are conducted through surrounding tissue. An experienced cardiologist can rapidly interpret a standard ECG tracing to diagnose a wide range of possible ailments[2]. However, because standard ECG traces only represent a short sampling of patient data, cardiac conditions which are irregular or intermittent may not be identifiable.

To address this shortcoming, many hospitals employ another class of ECG, called stress or Holter monitoring[1], to monitor patients in intensive care. Medical personnel deploy an ECG device that uses fewer electrodes (usually three or five), which provides a less detailed sampling of cardiac activity over an extended period. The cardiac rhythm is generally shown on a near-by display so that the patient's general cardiac condition can be monitored continuously. A physician may advise continuous monitoring if it is suspected that a patient has a cardiac problem, such as an irregular heartbeat, that occurs intermittently.

Many ECG machines, both standard and continuous, are marketed as "portable," but this does not always indicate that they are small and unobtrusive. Many such devices receive power from an electrical outlet and are sufficiently heavy that they must be mounted on a cart and wheeled from one location to the next. Wireless sensor networks have the potential to significantly improve this situation[1].

III. ECG ELECTRODES

Disposable sensor electrodes are used for testing purpose. They are pre-loaded with a low impedance wet Ag/Cl gel which is better than solid gels at reducing skin impedance, that is important for obtaining a clean signal. A tab connector allows for the cable to move around without disturbing the electrode-skin contact, which highly reduces the motion artifacts. Fig. 2 shows the picture of electrodes used for this purpose. We used these electrode for RA, LA and LL position. Three lead ECG signal is acquired for the wearable application.

Figure2. Disposable sensor electrode used for testing

ECG clamp electrodes can also be used for acquiring the ECG signal. Clamp electrodes with suitable universal connection are very suitable for getting ECG signal with better accuracy. The metallic surfaces of these electrodes are silver-silver chloride (Ag-AgCl) plated to provide maximum signal quality. Clamp electrodes are color coded IEC or AHA protocols (IEC color code: Red, Green, Yellow, Black and AHA color code: Red, Green, White, Black). Length of the clamp electrodes are 140 mm for adults.

Clamp electrodes are reusable electrodes and gel is used as the contact medium. ECG gel contains the free charges which guarantee electrical conduction. It is important to apply the electrode gel as it also minimizes the motion artifacts. Recent breakthroughs have been made in the form of insulated bio-electrodes (IBEs) as mentioned in [7]. They can measure the electric potential on the skin without resistive electrical contact and with very low capacitive coupling. This has been made possible by a combination of circuit design and the use of a new, low dielectric material. These IBEs enable through-clothing measurements.

IV.   HARDWARE DESIGN

Analog front-end processing is an important part of ECG signal processing, since it needs to distinguish between noise and the desired signal which is of small amplitude. The front-end processing circuitry consists of an instrumentation amplifier which reduces the common mode signal[9].
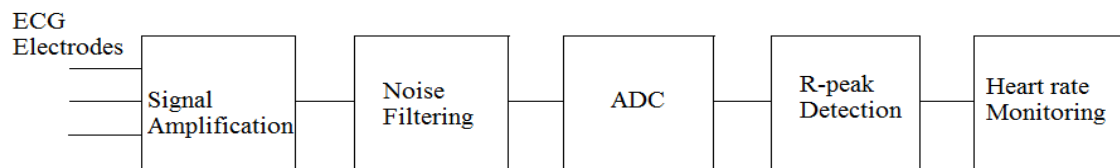
Figure3. Block diagram of wearable Heart-rate monitoring system

Based on information gathered by signal analysis, we can design specific procedures for signal processing. In order to enlarge the R-wave to about 0.5 – 1 V, the signal needs to be amplified by an amplifier (or several amplifiers) with a total gain of about 500. To remove the low-frequency noise, a high-pass filter can be used. The corner frequency of the filter should be between 0.03 to 0.1 Hz. Since the other two kinds of noise have frequency ranges that are overlapping with that of the ECG waveform, they are more difficulty to remove.

A set of buffer amplifiers is used before the instrumentation amplifier to acquire the Two lead ECG signal by grounding the third lead, as shown in the figure 3. This prevents the instrumentation amplifier from loading and faster acquisition of ECG data is possible. LM 324 is used for this purpose. It has wide bandwidth and large voltage gain. Also it works on very low power and hence is best suited for this circuit.
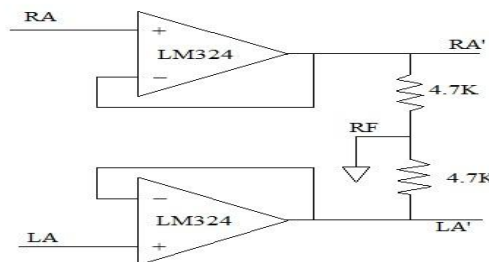
Figure4. Buffer amplifiers for ECG signal

The AD620, with its high accuracy of 10 ppm maximum nonlinearity, low offset voltage of 200μV max and offset drift of 2 μV/°C max, is ideal for use in precision data acquisition systems, such as weigh scales and transducer interfaces, Thermocouple amplifiers, Difference amplifiers, Low power data acquisition. Furthermore, the low noise, low input bias current, and low power of the AD620 make it well suited for medical applications such as ECG and noninvasive blood pressure monitors. Instrumentation amplifiers that operate on +/-6V are commonly used to take advantage of the large input voltage range. The instrumentation amplifiers should have high input impedance since the skin resistance could be very large. The gain of the instrumentation amplifier is kept near about 100. The signal chain for the ECG acquisition system consists of instrumentation amplifiers and filters.

In Fig. 5, different colors are used to help you to identify different stages in the overall signal processing. The circuit around AD620 with blue color corresponds to the block "1st Amplifier" in Fig. 1. The gain of the circuit is determined approximately by the following formula: $G = 50 K/ R_1$. The circuit with green color (C_1 and R_2) is the high-pass filter. The circuit around 741 with brown color is the 2nd Amplifier and the circuit with purple color (R_5 and C_2) is the low-pass filter.
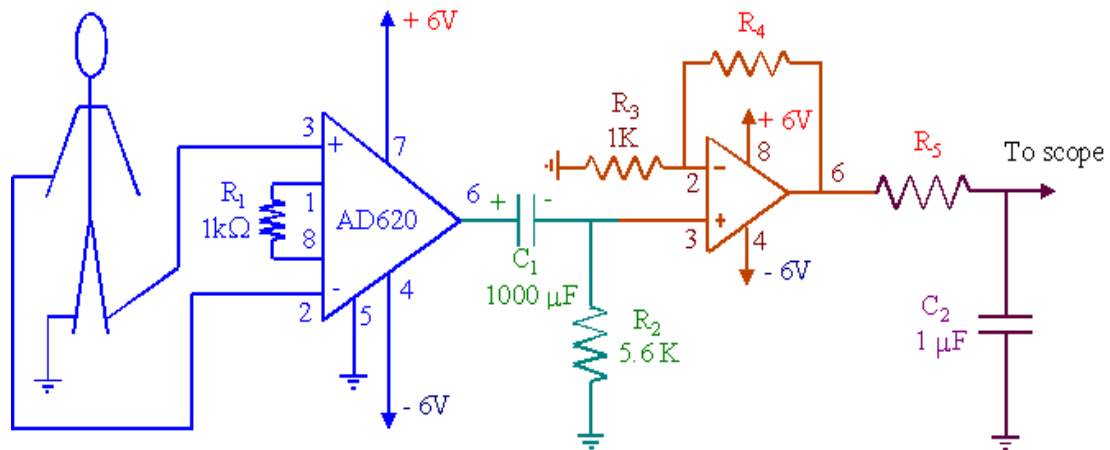
Figure5. Design of ECG signal processing scheme

Signal processing is a huge challenge since the actual signal value will be 0.5mV in an offset environment of 300mV. Other factors like AC power supply interference, RF interference from surgery equipment, and implanted devices like pace makers and physiological monitoring systems can also impact accuracy. The main sources of noise in ECG are[5]:

[1]   Baseline wander (low frequency noise)
[2]   Power line interference (50Hz or 60Hz noise from power lines)
[3]   Muscle noise (This noise is very difficult to remove as it is in the same region as the actual signal. It is usually corrected in software.)
[4]   Other interference (i.e., radio frequency noise from other equipment)

Baseline wander is a low frequency component present in the ECG system. This is caused due to offset voltages in the electrodes, respiration, and body movement. This can cause problems in the analysis of the ECG waveform. The offset also limits the maximum value of gain which can be obtained from the instrumentation amplifier. At higher gains, the signal can saturate. This noise can be removed by implementing a high pass filter. The cut-off frequency should be such that the ECG is undistorted while the baseline wander must be removed. A typical value of the cut-off frequency is 0.05Hz.

Most of the high frequency noise can be filtered before it is sampled by the ADC. The device can be shielded to prevent high frequency radiated noise from being coupled. So, a inbuilt low pass filter is used to remove the high frequency noise from the signal. The cut-off frequency for the low pass filter is 150Hz. This removes high frequency noises in the ECG signal.

PIC16F877A micro-controller is used for Heart rate monitoring. It contains 10 bit, 8 channel Analog-to-Digital Convertor (ADC).it consumes very less power and works on high speed FLASH/EEPROM technology. It has wide operating range. Fig. 5 shows the over-all flow-diagram of working of the controller.

The ECG signal is digitized and compared to the standard one. If the signal voltage is greater than the Standard voltage Vg, then the counter is incremented by one. After the given period of time is completed, the counter calculates the Heart beats. The heart beat count is multiplied with the desired multiplication factor and the heart rate is calculated. If the heart rate is beyond or above the standard limit, the alarm is given to the patient. Also there is provision to take serial data out. IN future, this data can be transmitted to the remote or nearest doctor using wired or wireless communication, so the doctor can regenerate the ECG wave whenever he wants to monitor the patient.

A graphical user interface is developed using Visual basic (software design). We can manually adjust the time duration of monitoring the ECG data in seconds. And a click on start button will start the Graphical recording of ECG on our GUI. The snapshot of the GUI is as shown in figure 7.
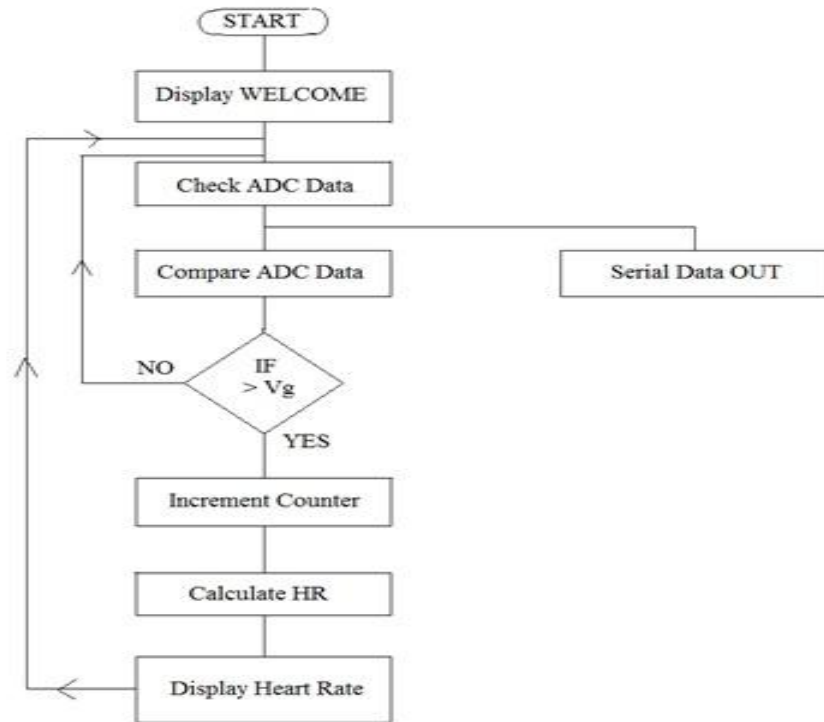
Figure6. Flow chart of working of PIC controller
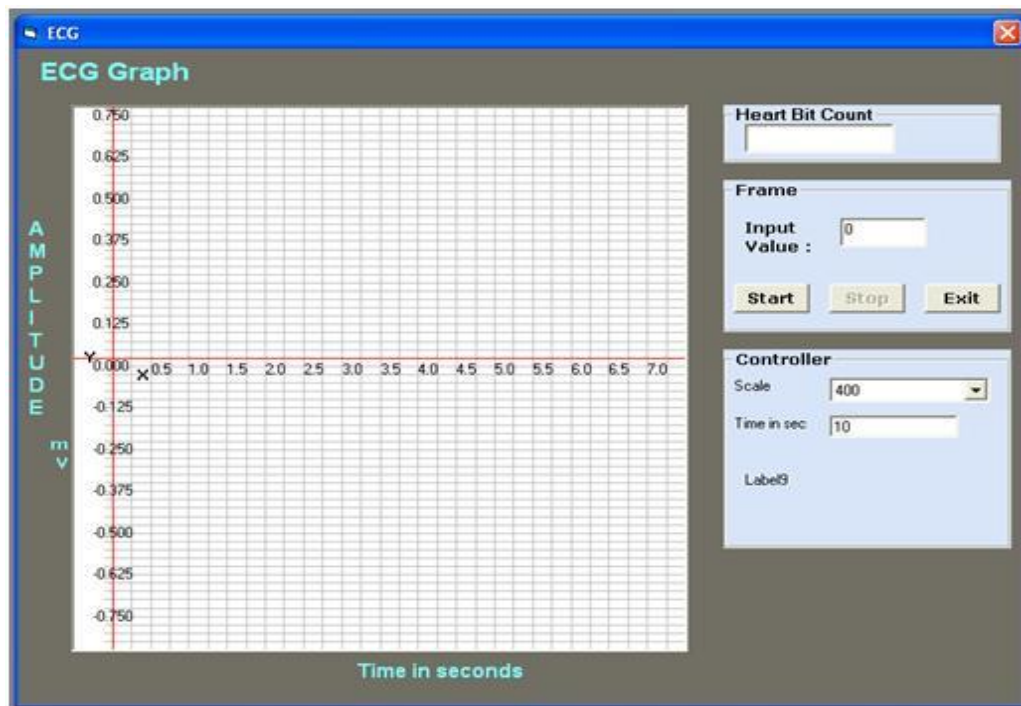
Figure7. Graphical User Interface for Wearable ECG Monitoring

V. RESULT

Figure 8 (a) and (b) are sample of results obtained form different patients. Results match with the doctors opinion as asked for the guidance. The results ware compared with the other techniques of heart rate monitoring. The performance is as better as other heart rate monitors used in hospitals. With this device more number of patients can also be monitored with some changes in the software.
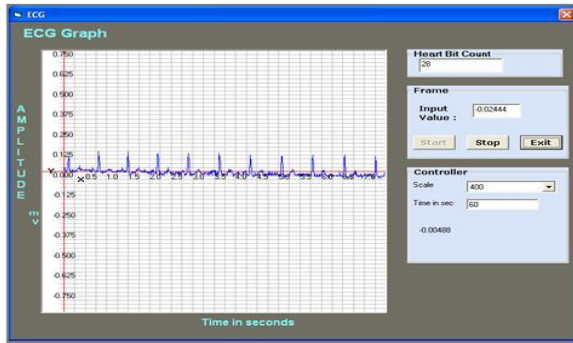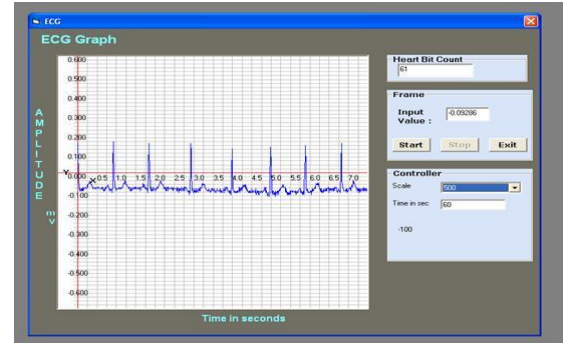
(a)   (b)

Figure8. (a) patient with heart-beat count 78 (normal) and (b) patient with heart-beat count 57 (Normal for post operative patient)

VI. CONCLUSION

In this paper, we have explained the signal processing of ECG signal to remove noise content and to obtain the well amplified signal. Also we discussed the algorithm of working of the microcontroller for heart beat calculation. Alarm is given for abnormal ECG and the patient can concern to the doctor. With the development of this device we have reduce the cost of ECG monitor. Also the device is wearable and can be easily operated by the patient anywhere he wants. Thus our basic aim to detect the emergency conditions before its too late is satisfied.

In Future, we can work on making some provision to transmit the Digital data on Hyper terminals to the doctor through wired or wireless communications, where the data will get stored at any specific location and doctor can regenerate the ECG wave from this data. Thus, the patient will receive good attention and this will reduce critical conditions. This helps in reducing cost and risks involved in personal Healthcare.

REFERENCES

[1] Steven A. Taylor, Hamid Sharif, "Wearable Patient Monitoring Application (ECG) using Wireless Sensor Networks" Proceedings of the 28th IEEE EMBS Annual International Conference New York City, USA, Aug 30-Sept 3, 2006.

[2] Wong Kiing-Ing, "A light-weighted, Low-cost and Wireless ECG Monitor Design based on TinyOS Operating System" 6th International Special Topic Conference on ITAB, Tokyo, 2007.

[3] C. J. Deepu, X. Y. Xu, X. D. Zou, Y. Lian, "An ECG-on-Chip for wearable Cardiac Monitoring Devices", IEEE Trans. 978-0-7695-3978-2/2010.

[4] B. U. Kohler, C. Hennig, and R. Orglmeister, "The principles of software QRS detection," IEEE Eng. Med. Biol. Mag., vol. 21, no. 1, pp. 42–57, Jan. 2002.

[5] Ajay Bharadwaj, Applications Engineer Sr, and Umanath Kamath, Contingent Workforce, Cypress Semiconductor Corp. "Accurate ECG Signal Processing".

[6] E. Katsiri, M. Ho, L. Wang, B. Lo and C. Toumazou, "Embedded Real-Time Heart Variability Analysis," IEEE Proceeding of the International Workshop on Wearable and Implantable Body Sensor Networks 2007.

[7] C. Park, P.H. Chou, Y. Bai, R. Matthews, A. Hibbs, "An ultra-wearable, wireless, low power ECG monitoring system", Biomedical Circuits and Systems Conference, Nov. 2006, pp. 241-244.

[8] S. Led, J. Fernandez and L. Serrano, "Design of a Wearable Device for ECG Continuous Monitoring Using Wireless Technology", International Conference of IEEE EMBS, San Francisco, CA, USA September 2004.

[9] Jerald Yoo, Long Yan, Seulki Lee, Hyejung Kim, and Hoi-Jun Yoo, "A Wearable ECG Acquisition System With Compact Planar-Fashionable Circuit Board-Based Shirt" IEEE transactions on information technology in biomedicine, vol. 13, no. 6, november, 2009

Survey on Energy-Efficient Secure Routing In Wireless Sensor Networks

Venkatesh Shankar[1] Dr Rajashree V Biradar[2]

CSE Dept SITCOE, Yadrav, India
CSE Dept BITM, Bellary, India

ABSTRACT

Data security and energy aware communication are key aspects in design of modern ad-hoc networks. In this paper we investigate issues associated with the development of secure IEEE 802.15.4 based wireless sensor networks (WSNs) – a special type of ad hoc networks. We focus on energy aware security architectures and protocols for use in WSNs. To give the motivation behind energy efficient secure networks, first, the security requirements of wireless sensor networks are presented and the relationships between network security and network lifetime limited by often in-sufficient resources of network nodes are explained. Second, a short literature survey of energy aware security solutions for use in WSNs is presented.

KEYWORDS: *energy aware security architectures, routing protocols, security protocols, wireless sensor networks, WSN.*

I. INTRODUCTION

Wireless Sensor Networks have emerged as an important new area in wireless technology. In the near future, the wireless sensor networks are expected to consist of thousands of inexpensive nodes, each having sensing capability with limited computational and communication power [1] , [2] and [3] which enable us to deploy a large-scale sensor network.A wireless network consisting of tiny devices which monitor physical or environmental conditions such as temperature, pressure, motion or pollutants etc. at different areas. Such sensor networks are expected to be widely deployed in a vast variety of environments for commercial, civil, and military applications such as surveillance, vehicle tracking, climate and habitat monitoring, intelligence, medical, and acoustic data gathering. The key limitations of wireless sensor networks are the storage, power and processing. These limitations and the specific architecture of sensor nodes call for energy efficient and secure communication protocols.Normally sensor nodes are spatially distributed throughout the region which has to be monitored they self-organize in to a network through wireless communication and collaborate with each other to accomplish the common task. Basic features of sensor networks are self-organizing capabilities, dynamic network topology, limited power, node failures and mobility of nodes, short-range broadcast communication and multi-hop routing, and large scale of deployment [4]. The strength of wireless sensor network lies in their flexibility and scalability. The capability of self-organize and wireless communication made them to be deployed in an ad-hoc fashion in remote or hazardous location without the need of any existing infrastructure. Through multi-hop communication a sensor node can communicate a far away node in the network. This allows the addition of sensor nodes in the network to expand the monitored area and hence proves its scalability and flexibility property.

The key challenge in sensor networks is to maximize the lifetime of sensor nodes due to the fact that it is not feasible to replace the batteries of thousands of sensor nodes. Therefore, computational operations of nodes and communication protocols must be made as energy efficient as possible. Among these protocols data transmission protocols have much more importance in terms of energy, since the energy required for data transmission takes 70 % of the total energy consumption of a wireless sensor network Area coverage and data aggregation techniques can greatly help conserve the scarce energy resources by eliminating data redundancy and minimizing the number of data transmissions. Therefore, data aggregation methods in sensor networks are extensively investigated in the literature [4], [5], [6] and [7].

Security in data communication is another important issue to be considered while designing wireless sensor networks, as wireless sensor networks may be deployed in hostile areas such as battle fields. Therefore, data aggregation protocols should work with the data communication security protocols, as any conflict between these protocols might create loopholes in network security.

II.   SENSOR NETWORK CHALLENGES

Wireless sensor network uses a wide variety of application and to impact these applications in real world environments, we need more efficient protocols and algorithms. Designing a new protocol or algorithm address some challenges are summarized below

2.1 Physical Resource Constraints:

The most important constraint imposed on sensor network is the limited battery power of sensor nodes. The effective lifetime of a sensor node is directly determined by its power supply. Hence lifetime of a sensor network is also determined by the power supply. Hence the energy consumption is main design issue of a protocol. Limited Computational power and memory size is another constraint that affects the amount of data that can be stored in individual sensor nodes. So the protocol should be simple and light-weighted. Communication delay in sensor network can be high due to limited communication channel shared by all nodes within each other's transmission range.

2.2 Ad-hoc Deployment:

Many applications are requires the ad-hoc deployment of sensor nodes in the specific area. Sensor nodes are randomly deployed over the region without any infrastructure and prior knowledge of topology. In such a situation, it is up to the nodes to identify its connectivity and distribution between the nodes. As an example, for event detection in a battle field the nodes typically would be dropped in to the enemy area from a plane.

2.3 Fault-Tolerance:

In a hostile environment, a sensor node may fail due to physical damage or lack of energy (power). If some nodes fail, the protocols that are working upon must accommodate these changes in the network. As an example, for routing or aggregation protocol, they must find suitable paths or aggregation point in case of these kinds of failures.

2.4 Scalability:

Most of the applications are needed the number of sensor nodes deployed must be in order of hundreds, thousands or more. The protocols must scalable enough to respond and operate with such large number of sensor nodes.

2.5 Quality of Service:

Some real time sensor application is very time critical which means the data should be delivered within a certain period of time from the moment it is sensed otherwise the data will be unusable So this must be a QOS parameter for some applications.

2.6 Security:

Security is very critical parameter in sensor networks, given some of the proposed applications. An effective compromise must be obtained between the low bandwidth requirements of sensor network applications and security demands for secure data communication in the sensor networks (which traditionally place considerable strain on resources)Thus, unlike traditional networks, where the focus is on maximizing channel throughput with secure transmission.

III.   SECURITY REQUIREMENTS OF WSN

Security for wireless sensor networks should focus on the protection of the data itself and the network connections between the nodes [8]–[10]. In general, security requirements often vary with application. In WSNs we can distinguish the following important requirements of security capabilities, authentication and authorization, availability, confidentiality, integrity and freshness. Thus, we need some mechanism for access authorization and protecting a mobile code. In many applications we need to protect fair access to communication channels and at the same time we often need to hide the information about physical location of our sensor node. Moreover, we need to secure routing and we have to defend our network against denial of service, malicious flows, node capturing and node injection, etc.

3.1 Authorization.

Data authorization specifies access rights to resources and is strongly related to access control. Access control should prevent unauthorized users from participating in network resources. Hence, only authorized users can join a given network. Access control relies on access policies that are formalized, like access control rules in a computer system. Most modern operating systems include access control.

3.2 Authentication.

Message authentication implies a sender verification using cryptographic key. Authentication mechanisms are used to detect maliciously or spoofed packets. They are especially important in WSNs which use a shared wireless medium. In case of unicast transmission, an authentication can be guaranteed by symmetric key cryptography, using Message Authentication Code (MAC) .

3.3 Availability.

In secure network data should be safe and accessible at all times. Availability guarantees the survivability of network services against Denial-of-Service (DoS) attacks that can be launched at any layer of a wireless sensor network, and may disable a given device (network node) permanently. Moreover, DoS attack involved excessive computation and communication may exhaust battery charge of a sensor device.

3.4 Confidentiality.

In WSN keeping sensitive data secret is the most important issue in case of critical applications in which highly sensitive data (secret keys, sensitive measurements, etc.) are collected and transmitted. Data confidentiality ensures that sensitive data is never disclosed to unauthorized users or entities. Hence, measurement data should not be available to neighboring nodes, and secure channels between nodes should be created. To protect a network against cyber attacks and malicious nodes, the routing information and sensor identities should remain confidential too. The standard approach to prevent end-to-end data confidentiality is to encrypt the data with a secret key.

3.5 Integrity and freshness

Data integrity is the quality of correctness, completeness, wholeness, soundness and compliance with the intention of the creators of the data. It is achieved by preventing unauthorized insertion, modification or destruction of data. In WSNs a malicious node may change messages to perturb the network functionality. Moreover, due to unreliable communication channels it is easy to inject infected packets or alerted data. In WSNs data integrity guarantees that a message being transferred is never corrupted, but providing data integrity is not enough for wireless communication. The compromised sensor nodes can listen to transmitted messages and replay attacks. Data freshness protects data against replay attacks by ensuring that the transmitted data is recent one.

IV. ENERGY EFFICIENT SECURITY ARCHITECTURES AND PROTOCOLS

4.1 SERP: Secure Energy Efficient Routing Protocol

The secure energy efficient routing protocol for wireless sensor networks (SERP) is described in [11]. The main idea of this protocol is to provide a robust transmission of authenticated and confidential data from the source sensor with limited energy budget to the base station. It is dedicated to WSNs with densely deployed relatively static sensor devices.

Three main objectives were considered during design of SERP:

➢ Energy aware organization of the network to ensure energy efficient transmission, and finally maximum lifetime of the network,
➢ Secure transmission: nodes should have the capability to detect falsely injected reports,
➢ Robust and resilient transmission: any node failure would not greatly hamper the performance of a network.

The protocol operates in two main phases: creating a back bone network and secure data transmission. A sink rooted tree structure is created as the backbone of the network taking into consideration balanced energy consumption. Next, a minimum number of forwarding nodes in the network is selected. The backbone network is restructured periodically. It is used for authenticated and encrypted data delivery from the source sensors to the base station. A one way hash chain and pre-stored shared secret keys are used for ensuring secure data transmission. An optional key refreshment mechanism that could be applied depending on the application is introduced for data freshness. The energy saving mechanism is based on disable the radio transceivers of selected nodes. The nodes in a network can operate in two main states *non-forwarding* – the transceiver is switched off, *forwarding* – both transceiver and sensing devices are switched on. It is assumed that after the backbone structure is constructed all nodes are either in forwarding or non-forwarding states.

4.2 EENC: Energy Efficiency Routing with Node Compromised Resistance

A novel energy efficiency routing protocol with node com- promised resistance (EENC) was developed by K. Lin *et al.*, and described in [12]. EENC bypasses the compromised nodes and improves the accuracy of packets under the condition of balancing the energy consumption. The Reinforcement learning based on the ant

Architecture	Security services	Properties
SPINS	Authentication, authenticated broadcast, confidentiality, integrity, freshness.	Consists of SNEP and µTesla (secure building blocks). Symmetric cryptography support. Encryption (CTR mode), Block Cipher (RC5). Not fully implemented and specified.
TinySec Authentication	Authentication, confidentiality, integrity, replay protection	Link layer architecture easily integrated into WSN. Symmetric cryptography support.

Table 1 : Summary of selected security architectures for WSN

colony optimization is used to complete routing tables. The trust values are assigned to all nodes of a network. The trust value is computed and based on the multiple behavior attributes such as packet drop rate, forwarding delay rate, etc. These values are used to detect the malicious nodes. Each node in a WSN computes the trust values of its one hop neighbors. The idea of EENC was to provide security with minimal energy consumption. To achieve this, each node storages trust values of all its neighbors and manages its energy resources.The EENC protocol operates as follows. To transmit data the secure and energy efficient route is computed. The calculation process consists of many rounds, each divided into three phases.• Routing detecting phase. A certain number of forward ants are generated to search for route leading to the sink. Each ant records the information about the minimum amount of energy and minimum trust value for nodes along the path, and the hop number for each node. • Pheromone updating phase. The sink node generates a backward ant, which carries all data collected by the forward ant. These data are used to update the pheromone value concerned with each node in a path. • Routing maintaining phase. The route for a given source and sink nodes is established based on trust values and updated pheromone values of the nodes carried during the pheromone updating phase. The Table 1 presents the summary of our survey – security architectures, provided services and their main properties

4.3.Summary and Conclusions

Many challenges arise from application of wireless ad hoc networking. We focused on one of them that is very important in wireless sensor networks – secure data protection and data transmission in WSN with limited resources. The paper provides a short overview of some representative energy efficient security techniques. We briefly discussed the security requirements of WSNs and showed the relationships between techniques for forming secure networks, and energy aware WSNs. Next, we described and compared based on literature survey selected energy aware architectures and protocols in WSNs that can be implemented in the physical, data link, network, and middleware layers of the OSI model.In summary, we can say that due to scarce resources, unique properties of wireless sensor networks, and often hostile environments it is a challenging task to protect sensitive information transmitted by nodes forming a WSN. Due to limited resources of nodes that form WSN many solutions providing strong security are impractical in this type of network. Therefore, we can find many security considerations that should be investigated in the nearest future.

REFERENCES

[1]. W. Su Y. Sankarasubramaniam E. Cayirci Akyildiz, I.F. A survey on sensor networks. *IEEE Communications Magazine*, 2002.

[2]. Kumar.S.P. Chee-Yee Chong. Sensor networks: Evolution, opportunities, and challenges. *Proc IEEE*, August 2003.

[3]. Ismail H. Kasimoglui Ian .F. Akyildiz. Wireless sensor and actor research challenges. *(Elsevier) Journal*, 2004.

[4]. D. Agrawal N. Shrivastava, C. Buragohain and S. Suri. Medians and beyond: new aggregation techniques for sensor networks. *Proceedings of the 2nd inter- national conference on Embedded networked sensor systems*, 2004. ACM Press.

[5]. Xiuli Ren and Haibin Yu1. Security mechanisms for wireless sensor networks. *IJCSNS International Journal of Computer Science and Network Security*, VOL.6(No.3):100-107, March 2006.

[6]. S. Setia S. Zhu and S. Jajodia. Leap: e±cient security mechanisms for large scale distributed sensor networks. *Proceedings of the 10th ACM conference on Computer and communications security*, 2003. ACM Press.

[7] P.Nair H.Cam, S.Ozdemir and D. Muthuavinashiappan. Espda Energy - efficient and secure pattern based data aggregation for wireless sensor networks. *Computer Communications IEEE Sensors*, 2006.

[8] M. Ahmad, M. Habib, and J. Muhammad, "Analysis of security protocols for Wireless Sensor Networks", in *Proc. 3rd Int. Conf. Comp. Res. Develop. ICCRD 2011*, Shanghai, China, 2011, vol. 2, pp. 383–387.

[9] C. Castelluccia, A. C.-F. Chan, E. Mykletun, and G. Tsudik, "Efficient and provably secure aggregation of encrypted data in wireless sensor networks", *J. ACM Trans. Sensor Netw. (TOSN)*, vol. 5, no. 3, 2009.

[10] S. R. Gandham, M. Dawande, R. Prakash, and S. Venkatesan, S., *Energy efficient schemes for wireless sensor networks with multiple mobile base stations*, in *Proc. IEEE Global Telecom. Conf. GLOBE- COM'03*, San Francisco, USA, 2003, vol. 1, pp. 377–381.

[11] A. K. Pathan and C. S. Hong, "SERP: secure energy-efficient routing protocol for densely deployed wireless sensor network", *Annales des Telecomm.*, pp. 529–541, 2008.

[12] K. Lin, Ch. F. Lai, X. Liu, and X. Guan, "Energy efficiency routing with node compromised resistance in wireless sensor networks", *Mob. Netw. Appl.*, vol. 17, pp. 75–89, 2012.

Flood Frequency Analysis of River Subernarekha, India, Using Gumbel's Extreme Value Distribution

Dr. Manas Kumar Mukherjee

[1,]*Associate Professor of Civil Engineering Jalpaiguri Government Engineering College*
West Bengal, India.

ABSTRACT

Estimation of Peak Flood Discharge for a desired return period is a pre-requisite for planning, design and management of hydraulic structures like barrages, dams, spillways, bridges etc. In this paper, a mathematical model has been developed between Peak Flood Discharge and Return Period using Gumbel's Extreme Value Distribution. The model will give reasonable estimate of Peak Flood discharge for any desired value of T, without any instrumentation and expensive and time consuming field work. Peak Discharge is a potential tool for designing important hydraulic structures like Concrete Gravity Dam, Weir, Barrage, and Bridge across the river, Guide bank etc. Moreover, the Stage corresponding to any given value of Peak Discharge can readily be ascertained by developing Rating Curves following the procedure given by the Author as referenced below. This Stage will be helpful in maintaining Danger Level Flood of the river Subernarekha. Emergency evacuation may be adopted by propagating well advanced 'Flood Warning' that may save thousands of lives from the fury of flood, may be put in place.

KEY WORDS*: River Subernarekha, Peak Flood Discharge, Return Period, Gumbel's Extreme Value Distribution, Confidence Limit, Stage. Chi-Square test.*

I. INTRODUCTION:

Estimation of Peak Flood Magnitude for a desired return period is often required for planning, design and management of hydraulic and other structures in a region. These events are essential in the post commissioning stage where in the assessment of failure of hydraulic structures needs to be carried out. (Wyno Journal of Engineering & Technology Research, Vol. 1(1) PP- 1-9 March, 2013). In this paper, 6-h unit hydrograph data is used for Peak Flood Estimation to arrive at a design parameter for a region. In extreme value theory, probability distribution of Gumbel is widely used for frequency analysis of recorded meteorological data such as rainfall, temperature, wind speed, evaporation, Peak Flood etc and hence used in the present studyThe Subernarekha [http:\\www.springerlink.com/ content/7885062173413017/] is an inter-state river flowing through Bihar, West Bengal and Orissa states. It starts in the Chotanagpur Plateau of Bihar and flows into the Bay of Bengal. The upper part of the Subernarekha and its tributaries run through the fertile land of Bihar, but the farming in this region mainly depends on the inadequate and ultimate rains, and the water resources of the Subernarekha river system remain largely untapped. The upper basin, besides containing fertile land, also contains large reserves of minerals. A number of important industries have therefore grown along the banks of the river.

Subernarekha Multipurpose Project is a joint project of Jharkhand, Orissa and West Bengal state for which tripartite agreement was signed between undivided Bihar, Orissa and West Bengal on 07.08.1978. The benefits of the project is furnished hereunder-

a) JHARKHAND:
   i) The creation of Irrigation Potential - 2, 36,846 ha.
   ii) Municipal and Industrial Use   - 740 MCM annually
   iii) Hydel Power Production - 8 MW
b) ORISSA:
   i) Creation of Irrigation Potential - 90,000 ha
   c) WEST BENGAL:
   i) Creation of Irrigation Potential - 5,000 ha

Catchment characteristics such as, stream order, drainage density, stream density, length, shape, slope, etc., [Reddy, J, R., 1998] and Annual Peak Flood Magnitude were not available. Instead, one 6-h unit

hydrograph for Kharkai Barrage Site was used for the present study (Data Source: Irrigation International Building, Salt Lake City, Kolkata, Government of West Bengal, India).

II.   PROCESSING OF THE COMPUTER OUTPUT DATA

By using the method of superposition, the unit hydrographs of different durations have been obtained. In this method, if a D-hour unit hydrograph is available, and it is desired to develop a unit hydrograph of nD-hour duration, where n is an integer, it is easily accomplished by superposing n unit hydrographs with each graph separated from the previous one by D-hour [Subramanya, K, 1994]. A Computer Program has been developed for this purpose. First of all, from the computer output, the unit hydrograph for each duration of Kharkai Catchment has been developed. Then from the unit hydrograph thus developed, the Peak Discharge (Q_p) and corresponding D have been identified. Before use, the data has been statistically checked for consistency and continuity.

III.   GUMBEL DISTRIBUTION

Gumbel distribution is a statistical method often used for predicting extreme hydrological events such as floods (Zelenhasic, 1970; Haan, 1977; Shaw, 1983). In this study it has been applied for flood frequency analysis because (a) peak flow data are homogeneous and independent hence lack long-term trends; (b) the river is less regulated, hence is not significantly affected by reservoir operations, diversions or urbanization; and (c) flow data cover a relatively long record and is of good quality (Mujere, 2006). (IJCSE-ISSN: 0975-3397, Vol. 3 No. 7 July 2011). The equation for fitting the Gumbel distribution to observed series of flood flows at different return periods T is (Sarma, 1999).

$$X_T = \overline{X} + K.\sigma_{n-1} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \quad (1)$$

Where, X_T denotes the magnitude of the T- year flood event, K is the frequency factor; ẋ and σ_{n-1} are the mean and the standard deviation of the maximum instantaneous flows respectively.

The reduced variate is expressed as $\quad y_T = -[\ln.\ln.(T/T-1)]\dots\dots \quad (2)$

The frequency factor expresses as $K = (y_T - y_n)/s_n \dots\dots\dots\dots\dots\dots.. \quad (3)$

Where, y_n and s_n are taken from, standard Table of (K Subramanya, 2004).

The Chi-square ($\chi 2$) test was carried out to find goodness of fit between the measured and predicted flood flows. It was applied to test the hypothesis that the flood data fit Gumbel distribution. Details of Chi-square ($\chi 2$) test has been furnished in Table-4 her under.

The detailed Computation table and the model in graphical & equation form are furnished here under.

Table showing computation details by Gumbel's Extreme – Value distribution

Table-1

Peak Flood Discharge (Qp) in Cumec (Field Value)	Sample Size, N	Mean of Series, $\overline{X}$, in Cumec	STDEV of the Series, $\sigma_{(N-1)}$, in Cumec	Return period T in Years	Reduced Variate, y_T	Frequency Factor, K	Computed Peak Flood Discharge X_T in Cumec obtained by Gumbel's method
862.00	100	104.93	140.0465	2.5	0.67172	0.09	117.53
657.50				5	1.49999	0.78	214.17
574.67				7.5	1.94420	1.14	264.58
502.50				10	2.25037	1.40	300.99
439.80				12.5	2.48432	1.59	327.60
388.00				15	2.67375	1.75	350.01
346.14				17.5	2.83292	1.88	368.21
313.37				20	2.97019	1.99	383.62
284.22				22.5	3.09087	20.9	397.62
258.80				25	3.19853	2.18	410.23
236.10				27.5	3.29572	2.40	421.43
216.42				30	3.38429	2.34	432.64
199.77				32.5	3.46565	2.40	4410.4
185.50				35	3.54088	2.47	45.084
173.13				37.5	3.61085	2.52	457.84
162.31				40	3.67624	2.58	466.25
152.76				42.5	3.73762	2.63	473.25

144.28				45	3.7945	2.68	480.25
136.68				47.5	3.85010	2.72	485.85
129.85				50	3.90193	2.77	491.45
123.67				52.5	3.95121	2.81	498.46
118.04				55	3.99817	2.85	504.06
112.91				57.5	4.04302	2.88	508.26
108.20				60	4.08595	2.92	513.86
103.88				62.5	4.12711	2.95	518.06
99.88				65	4.16664	2.98	522.26
96.18				67.5	4.20467	3.02	527.87
92.75				70	4.24131	3.05	532.07
89.55				72.5	4.27665	3.08	536.27
86.57				75	4.31078	3.10	539.07
83.77				77.5	4.34379	3.13	543.27
81.16				80	4.37544	3.16	547.47
78.70				82.5	4.40670	3.18	550.27
76.38				85	4.43674	3.21	554.47
75.20				87.5	4.46589	3.23	557.28
72.13				90	4.49422	3.26	561.48
70.19				92.5	4.52177	3.28	564.28
68.34				95	4.54859	3.30	567.08
66.59				97.5	4.57470	3.32	569.88
64.92				100	4.60014	3.34	572.68

Developed model is furnished here under:

PLOT OF PEAK FLOOD DISCHARGE (Xт) & RETURN PERIOD (T) FOR RIVER SUBARNAREKHA, INDIA USING GUMBEL'S EXTREME-VALUE DISTRIBUTION

$X_T = 104.93 + 140.0465K$
$K = f(T)$
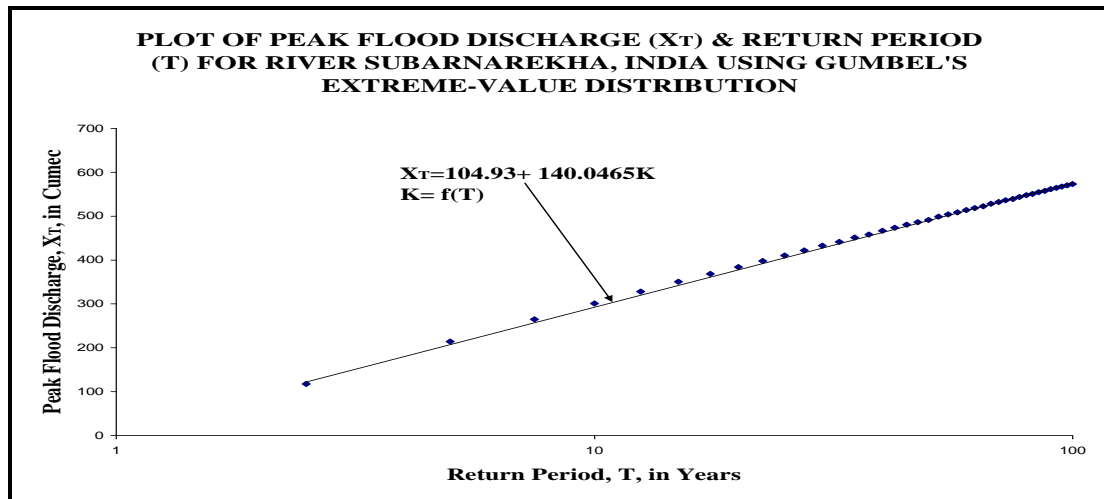
Figure-1

Confidence limit:

Table:2

Sample N	Mean of Series	STDEV of Series	Return period T in Years	Frequency factor K	Upper bound value of Peak flood discharge, X2 in Cumec	Peak flood discharge, X-T computed by Gumbel's method in Cumec	Lower bound value of Peak flood discharge, X-1 in Cumec
100	104.93	140.0465	2.5	0.09	146.643	117.53	88.416
			5	0.78	287.80	214.17	140.53
			7.5	1.14	371.91	264.58	157.25
			10	1.40	437.53	300.99	164.45
			12.5	1.59	488.06	327.60	167.137
			15	1.75	517.88	350.01	167.705
			17.5	1.88	569.88	368.21	167.54

20	1.99	601.57	383.62	165.61
22.5	2.09	631.49	397.62	163.75
25	2.18	658.86	410.23	161.59
27.5	2.26	683.65	421.43	159.21
30	2.34	708.68	432.64	156.59
32.5	2.40	727.95	441.04	154.13
35	2.47	750.524	450.84	151.156
37.5	2.52	766.84	457.84	148.84
40	2.58	786.64	466.25	145.86
42.5	2.63	806.01	473.25	140.48
45	2.68	820.06	480.25	140.43
47.5	2.72	833.85	485.85	138.07
50	2.77	849.29	491.45	133.60
52.5	2.81	864.50	498.46	132.41
55	2.85	878.34	504.06	129.78
57.5	2.88	888.57	508.26	127.94
60	2.92	902.822	513.86	124.90
62.5	2.95	913.196	518.06	122.924
65	2.98	923.98	522.26	120.54
67.5	3.02	938.37	527.87	117.36
70	3.05	949.07	532.07	115.06
72.5	3.08	959.94	536.27	112.59
75	3.10	967.16	539.07	110.98
77.5	3.13	978.05	543.27	108.48
80	3.16	989.03	547.47	105.90
82.5	3.18	996.36	550.27	104.18
85	3.21	1007.23	554.47	101.71
87.5	3.23	1014.84	557.28	99.718
90	3.26	1026.04	561.48	96.92
92.5	3.28	1033.504	564.28	95.056
95	3.30	1040.96	567.08	93.20
97.5	3.32	1048.43	569.88	91.326
100	3.34	1056.10	572.68	89.66
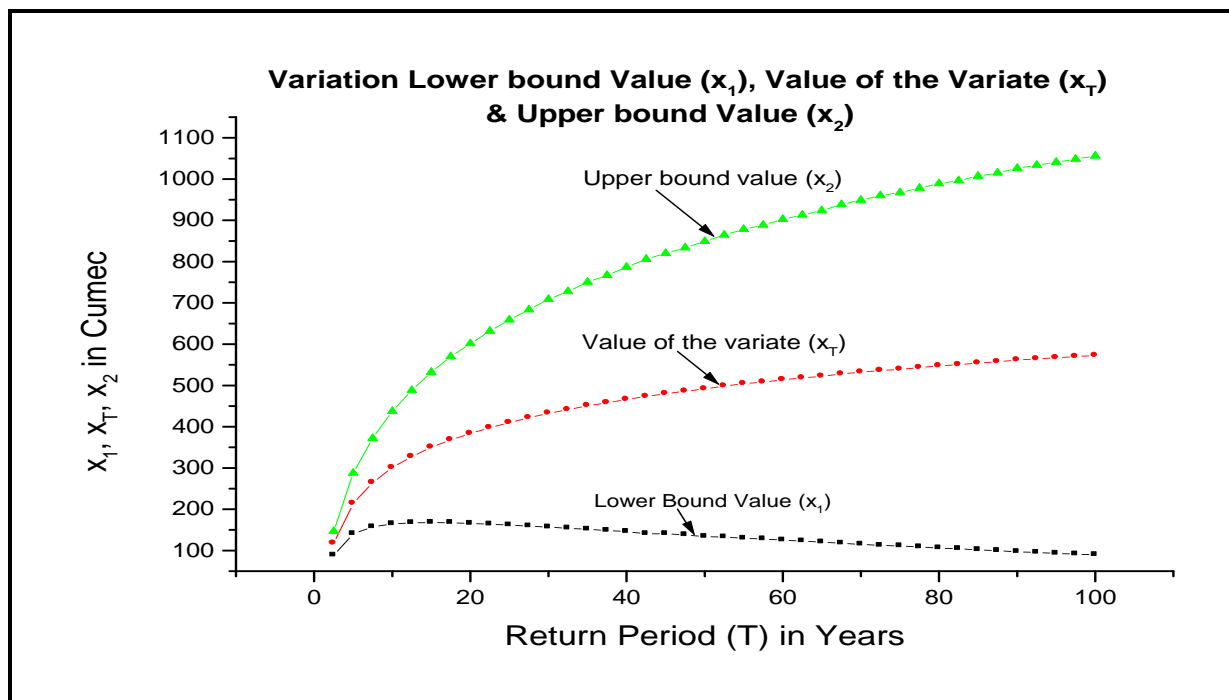
Figure-2

Table Showing Comparison of Peak Flood Discharge Computation by Gumbel's Extreme Value Distribution and Empirical Model developed by Author Referenced in the following Table 3

Table-3

Return Periods T in Years	Peak Flood Discharge X_T (Cumec) Computed by Gumbel's Method	Peak Flood Discharge Qp (Cumec) Computed by Empirical Model developed by the Author(IJCR Vol-4, Issue,04, pp-164, April, 2012)	% Deviation	Absolute % Deviation	Average of Absolute % Deviation	Remarks
2.5	117.53	66.12	43.74	43.74		Hence, it can be concluded that the Peak Flood Discharge computed by the aforesaid two methods are reasonably close to each other. But, obviously model obtained from Gumbel's Method is more accurate to some extent because it is a pure statistical extreme-
5	214.17	128.68	39.86	39.86		
7.5	264.58	186.10	29.66	29.66		
10	300.99	238.60	20.72	20.72		
12.5	327.60	286.39	12.57	12.57		
15	350.01	329.68	5.83	5.83		
17.5	368.21	368.71	-0.13	0.13		
20	383.62	403.67	-5.22	5.22		
22.5	397.62	411.49	-3.488	3.488		
25	410.23	462.27	-12.68	12.68		
27.5	421.43	486.34	-15.40	15.40		
30	432.64	507.22	-17.23	17.23	14.75	
32.5	441.04	525.11	-19.06	19.06		
35	450.84	540.23	-19.83	19.83		
37.5	457.84	552.81	-20.74	20.74		
40	466.25	563.05	-20.76	20.76		
42.5	473.25	571.17	-20.70	20.70		
45	480.25	577.38	-20.22	20.22		
47.5	485.85	581.90	-20.93	20.93		
50	491.45	584.96	-19.02	19.02		
52.5	498.46	586.75	-17.70	17.70		
55	504.06	587.50	-16.55	16.55		
57.5	508.26	587.43	-15.58	15.58		
60	513.86	586.75	-14.18	14.18		

62.5	518.06	585.57	-13.03	13.03		value
65	522.26	584.41	-11.9	11.9		distribution
67.5	527.87	583.79	-10.48	10.48		method.
70	532.07	582.22	-9.49	9.49		
72.5	536.27	581.71	-8.47	8.47		
75	539.07	581.89	-7.94	7.94		
77.5	543.27	582.97	-7.3	7.3		
80	547.47	585.17	-6.88	6.88		
82.5	550.27	588.69	-6.98	6.98		
85	554.47	593.76	-7.08	7.08		
87.5	557.28	600.59	-7.77	7.77		
90	561.48	609.40	-8.53	8.53		
92.5	564.28	620.40	-9.94	9.94		
95	567.08	633.80	-11.76	11.76		
97.5	569.88	649.84	-14.03	14.03		
100	572.68	668.71	-16.77	16.77		

Goodness of fit:

Table – 4

Summary of $\chi 2$ Test Results for the river Subernarekha

River	$\chi 2$ (Computed)	D.O.F	$\chi 2$ from Table		Remarks
			95% Confidence	99% Confidence	
Subernarekha	9.219	4	9.49	13.28	Passed at 95 % and 99% confidence

(Source: $\chi 2$ from Table: N. G. Das, 1996; Table-I & II)

IV.   DISCUSSION

[1]   It has already been established that Value of the Variate X_T is unbounded. Figure-2 shown above strongly supports this statement. Here, variation of X_1, X_T and X_2 with T are truly convergent in nature.

[2]   Moreover, the author has developed an empirical model between Peak Flood Discharge (Q_p) .vs. Return Period (T), (IJCR, Vol-4, Issue, 04, pp-164, April, 2012). That empirical model has been compared with the model developed here by Gumbel's method and the comparison has been furnished in Table-3.

[3]   It has been observed that the Peak Flood Discharge for a given Return Period (T) computed by two models mentioned above do not vary too much.

[4]   4.   For a given Return Period (T), Peak Discharge can be computed by any of the two models, particularly at higher values of Return Period (T).

V. CONCLUSION

- For any anticipated T, X_T can readily be estimated from the developed model as shown in the Figure-1 and corresponding equation has also been furnished there.
- However, the model will give reasonable estimate of X_T for any desired value of T, without any instrumentation and expensive and time consuming field work.
- For any anticipated value of T, X_T can readily be ascertained from the developed model suggested above and the Stage (G), corresponding to X_T can be estimated following the procedure as given by [Mukherjee, M,K., and Sarkar, S., 2007].
- These Stages may be obtained from Stage-Discharge (G-Q) model, corresponding to X_T Therefore, the values of G thus obtained are on conservative side.
- If presently adopted Danger level for 'Flood' for the river Subernarekha at the gauging site, is lower than the stage computed from (G-Q) model, then there is no problem.
- If presently adopted Danger level for 'Flood' for the river Subernarekha at the gauging site, is higher than the stage computed from (G-Q) model, then the presently adopted danger level for flood needed to be changed.
- Therefore, emergency evacuation may be adopted by propagating well advanced 'Flood Warning' that may save thousands of lives from the fury of flood, may be put in place.
- 'Flood Plain Zoning' may also be introduced to protect the lives of thousands of people and their properties, to minimize the socio-economic disaster created by flood.
- Moreover, Peak Discharge (X_T) is a potential tool for designing important hydraulic structures like Concrete Gravity Dam, Weir, Barrage, Bridge across the river, Guide bank etc.
- The entire water resource of river Subernarekha is largely untapped. Hence, construction of hydraulic structures will be helpful for resource generation also.

IV. NOTATIONS USED IN THE PAPER

- **G** - Stage
- **Q** - Discharge
- **Q$_p$** - Peak Discharge of Unit Hydrograph
- **T** - Return Period
- **X$_T$** - Gumbel's Variate for a return period T
- $\overline{X}$ - Mean of the series
- **N** - Sample Size
- **P** - Probability of an event equaled to or exceeded
- **STDEV** - Standard Deviation of percentage deviation
- **X$_1$** - Lower bound value of X$_T$
- **X$_2$** - Upper bound value of X$_T$
- $\chi 2$ - Chi-Square

REFERENCES

[1]. Chow. V. T., Maidment. David R. and Mays. Larry W.1988. Applied hydrology; McGraw-Hill, New York, USA, 1988

[2]. Beard. L. R. Statisitical analysis in hydrology. Trans. Amer. Soc. Civil. Eng.,108,1110-1160, 1943.

[3]. Charles. T. Haan. Statistical Methods in Hydrology. Ewp Affiliated East-West press Iowa State University Press, 1995.

[4]. Chow. V. T. 1964. Handbook of Applied Hydrology. McGraw-Hill. New York. USA, 1964.

[5]. Das. N.G. 1996. Statistical Methods in Commerce. Accountancy & Economics (Part-II). M. Das & Co. Kolkata, 1996.

[6]. http:\\www.springerlink.com/ content/7885062173413017/.

[7]. Vivekanandan. N. Wyno Journal of Engineering & Technology Research. Vol. 1(1). PP. 1-9 March, 2013.

[8]. Mujere. Never. International Journal on Computer Science and Engineering(IJCSE). ISSN: 0975-3397. Vol. 3 No. 7, July 2011.

[9]. Irwin. Miller. John. and Freund. Probability and Statistics for Engineers. Prentice Hall of India Private Limited. New Delhi, 1985.

[10]. Levin. Richard. Statistics for Management. Prentice Hall of India Private Limited. New Delhi-1, 1986.

[11]. Mukherjee. M. K. and Sarkar. S. Deterministic Modeling of Stage-Discharge Relationship of the River Sankosh North Bengal. Institute of Landscape, Ecology & Ekistics. Vol.30(I). Page-171-176, 2007.

[12]. Mukherjee. M. K. Studies on Empirical model between Discharge (Q$_p$) and Return period (T) of Major Himalayan Rivers in North Bengal, National Conference on Integrated Water & Wastewater Management at School of Water Resources Engineering. Jadavpur University. Kolkata-700032, 2008.

[13]. Mukherjee. M. K. Studies on Empirical modeling between Annual Peak Discharge (Qp) and Return Period (T) of River Sankosh in North Bengal using Plotting Position Formulae. Institute of Landscape, Ecology & Ekistics. Vol.30(1). Page-358-367, 2009.

[14]. Mukherjee. M. K. Development of Hydrological Models between Peak Discharge (Q$_p$) .vs. Duration (D) of Unit-Hydrograph and Peak Discharge (Q$_p$) .vs. Return Period (T) at Galudhi Dam Site and Kharkai Barrage Site, India. AMSE. France. Paper No-10573(2C), 2011.

[15]. Mukherjee. M. K. Model of Peak Discharge (Q$_p$) & Return Period (T) of river Subernarekha, India. International Journal of Current Research. Vol-4. Issue 04. pp-164, April, 2012.

[16]. Reddy. J. R. 1988. A text book of Hydrology; Laxmi Publishers, 1988.

[17]. Subramanya. K. 1994. Engineering Hydrology. Tata McGraw-Hill Publishing Company Limited. New Delhi, 2004.

ACKNOWLEDGEMENT

Measures To Improve the Declining Usage and Operation of School Farm in Secondary Schools in Ekiti State, Nigeria

BY

FAMIWOLE, Remigius O. (Ph.D)
Department Of Curriculum Studies
Ekiti State University, Ado-Ekiti, Nigeria

ABSTRACT

It has been observed that only few secondary schools have operational standard school farms, where practical agriculture can be learnt to complement classroom instructions in secondary schools. The main focus of this paper was to evolve the measures that can be adopted to improve the declining usage and operation of standard school farm in all secondary schools in Ekiti state. The study was a descriptive survey research. The population used comprised all the 602 teachers of agricultural science in both public and private secondary schools in the States. A combination of purposive, stratified and simple random sampling techniques were uesed to select the 160 teachers of agriculture who responded to the 4 sets of questionnaires used. The constructed questionnaires were face and content validated. A test retest method was used and a reliability co-efficient of 0.86 was obtained using Pearson Product Moment Co-relation Co-efficient formula. Frequency counts, percentages, ranking order, mean, standard deviation and t value were used to analyze the data used for the study, and test the formulated hypothesis respectively. The findings of the study revealed, among others, the major causes of the decline and the measures that can be adopted to improve the usage and operation of school farms. The study recommended that the Ministry of Education should give specifications for standard school farms; the Ministry in collaboration with relevant agencies should supervise and assist in funding supply of tools and implements for managing school farm in Ekiti State.

KEY WORDS: *Learning experience, skill development, learning by doing, practical agriculture, crop rotation.*

I. INTRODUCTION

The school farm is a laboratory, specifically designed and operated, for the purpose of carrying out practicals in agricultural science or education in order to impart knowledge and managerial skills to students through practice. It is an area specifically earmarked for agricultural activities, usually sited in the school or at a walking distance to the school compound. Olaitan (2001), in a study noted that students acquire agricultural knowledge in classrooms in such areas like crop production, forestry, fish farming, agricultural business, farm management, livestock production and so on. He stated that the school farm is an agricultural laboratory that interprets the acquired theoretical knowledge into practice through practical activities to gain experience.
Students are guided on standard school farms through hands on experience to put the knowledge gained in the classrooms into practice on the school farm under the watch of the teacher. The teacher makes use of different methods and technologies, such as demonstration, observation, imitation, and supervised practice to explain techniques and complement the different learning experience required to teach the practicals very effectively.

The school farm helps to inculcate into the students the need to value what they could not do or practice by themselves and for themselves. The school farm encourages the use of the head to think, eyes to see and to lay hands on actual operation or techniques to make learning easier and more permanent. According to Mama (2001), students are made to use their heads, hands and hearts (3Hs) during practical activities on the school farm to produce crops, livestock, keep records and participate in managerial activities. The respect for dignity of human labour is also built into the learners through this process. They also learn how to protect the farm from pests, and diseases, predators, human invaders and how to store farm produce and process them. Effective use of the school farm allows for skill development among the growing youths, and transfer of

knowledge or experience. For example, students from different background, especially those from farming families with some learning experience in agricultural activities, through exposure to school farm work, would see the relevance of their experience to the agricultural activities of the school and vice versa, thereby encouraging the transfer of modern ways of farming from their schools to the rural farmers for improvement purpose.In 2008 the National Education Research Council (NERDC), stipulated that the Agricultural Science Curriculum is set up to ensure the acquisition of productive (entrepreneurial) work skills through prescribed activities and projects which are inherent aspects of the applied technology called agriculture. NERDC specified that for effective delivery of the subject matter, improved achievement by the learners, only subject specialist teachers are to be engaged in the implementation of the curriculum. According to the body, there should be teacher orientation, training and re-training for effective delivery. The school are to provide necessary logistic for the successful implementation of the whole agricultural science curriculum; while the school farm should be seen as a field or laboratory for the training of the Basic Education Learners, with the basic focus on skill development, and self reliance.

II.   THE PROBLEM AREA

The National Curriculum for Secondary School (2008) recommended that all secondary schools must keep a standard school farm where crops are to be grown with at least one species of livestock from either pigs, rabbits, poultry and goats, sheep, cattle and where feasible, fish production. It also emphasized the use of practical note books, weed albums, insect box, record books to be used by individual students. The syllabus provides that the school farm should be used for training students in skill development through demonstration, observation and practice. One would expect the final examination in the subject to involve demonstration of acquired skills by students, but the examination covers mainly the theory and alternate to practical. Nowadys students who can read textbook but have never sighted a growing maize plant or a farm could pass both internal and external examinations in agricultural science creditably. This is observed to be unethical to national development and capacity building.Olaitan and Mama (2001) observed that in most secondary schools, the farms are not well developed and utilized to develop appropriate skills and interest in students. According to them most students are passing such examinations like Senior School Certificate Examination (SSCE) or West African School Certificate, National Examination Council Exams or even NABTEB in Agricultural Science without sighting a farm. The researcher observed that there are only few school farms in some secondary schools in Ekiti State Nigeria. Where some exist they are either not well managed or rather not effectively used for teaching, skill development or learning purpose. Similarly, most students studying agricultural science have never touched a bird, they have never keep farm records, cannot apply fertilizer, spray chemicals, or manage a farm. According to Olaitan (2010), they lack farm ethics, yet they record success at internal and external examinations in agricultural science. Most students passing out of secondary schools nowadays are "robots". The situation according to Olaitan would continue to increase the unemployment graph and complicate the problems of youth delinquency because the students are not marketable. They acquire no saleable skills while in school. This study was designed, therefore, to identify the causes of decline in the use and operation of school farm to development saleable skills in the students and for effective teaching and learning of agricultural concepts.

2.1.Research Questions
[1]   The following questions were raised and answered in this study.
[2]   What are the causes of decline in the use of school farm for skill development in secondary schools?
[3]   What are the measures that can be adopted to improve the use of school farm for skill development in agriculture?
[4]   What are the trends in the use and operation of school farm for skill development in secondary schools?

2.2.Hypothesis
The study will test this hypothesis at $P > 0.05$ level of significance.
H_0:   There is no significant difference in the perception of respondents (teachers in public and private) secondary schools in the declining nature of school farm usage and operation. For skill development in secondary schools in Ekiti State.

2.3.Methodology
The study was a descriptive survey method. The population used comprised all the teachers teaching Agricultural Science in both private and public secondary schools in Ekiti State, Nigeria, tolalling 602, (Ekiti Ministry Education, 2012). Stratified random sampling technique was used to select 162 teachers used for the study. The 162 respondents were further stratified on the basis of public (128) and private schools (64) , that is

eight of public and four teachers in private secondary schools in each of the sixteen Local Government Area of the state

2.4. Data Collection

Four structured questionnaire were used for data collection. The instrument were face and content validated. A test-retest method was used, a reliability co-efficient of 0.86 was obtained using Pearson Production Moment Correlation Coefficient formula.

2.5. Data Collection and Analysis

The questionnaire were administered on the respondents by personal contact. Three research assistants were engaged to personally administer the questionnaire by hand on the respondents, and collect back after completion. Out of the 162 questionnaire distributed, 150 copies were return and dully completed representing a 78.13 percentage return rate.

2.6. Data Analysis

Colleted data were analysed using frequency distribution, percentage, ranking order, mean, standard deviation and t-test was used to test the hypothesis at $P < 0.05$ level of significance

Table 1: Number of Agric Teachers in Public and Private Secondary Schools in Ekiti State.

S/N	Local Government Area	Public	Private	Total
1.	Ado-Ekiti	66	46	110
2.	Efon Alaaye	11	02	13
3.	Ekiti East	27	13	40
4.	Ekiti South West	29	05	34
5.	Ekiti West	30	04	34
6.	Emure	12	04	16
7.	Gbonyin	26	07	33
8.	Ido/Osi	35	06	41
9.	Ijero	26	07	33
10.	Ikere	38	07	45
11.	Ikole	35	09	44
12.	Ilejemeje	11	02	13
13.	Ile/Ife	40	04	44
14.	Ise/Orun	19	12	31
15.	Moba	23	06	29
16.	Oye	34	08	42

Source: (Ekiti State Teaching Service Commission, 2012)

2.7. Research Question 1: What are the causes of decline in the use and operation of school farms for skill development in agricultural science in secondary school?

Table 2: Frequency counts and ranking order percentage of causes of decline in the use and operation of school farm for skill development

S/N	Causes of decline in the use of school farm	YES			NO		REMARKS
		NO	%		NO	%	
1.	Non-relevance of school farm to pass agric examinations at all levels.	120	80	7th	30	20	YES
2.	Insufficient land to establish school farm	23	15	19th	127	85	NO
3.	Lack of skilled labour for practical agriculture.	80	53	13th	70	47	YES
4.	Lack of interest in practical on the part of the students.	105	70	10th	45	30	YES
5.	Congested time table not in favour of practical agriculture.	128	85.5	8th	22	14.7	YES

6.	Lack of interest in practical agric on the part of the school management.	88	59	12th	62	41	YES
7.	Shortage of trained staff to teach practical agriculture.	23	9	21st	137	91	NO
8.	Students immaturity for practical agriculture.	49	26	17th	111	74	NO
9.	Poor funding and finance of practical agriculture.	123	82	6th	37	18	YES
10.	Lack of tools and simple farm implements.	69	46	16th	81	54	NO
11.	Poor climatic and environmental condition.	28	12	20th	132	88	NO
12.	No supervision or monitoring of farm activities by the Ministry of Education.	107	71	10th	73	29	YES
13.	Students lack of interest in the agric teachers handing farm work	99	66	11th	51	34	YES
14.	Un-coperative altitude of staff and management.	110	73	9th	40	27	YES
15.	The use of other teachers to teach agriculture	113	75	8th	47	25	YES
16.	Practical agriculture not made interesting/motivation	129	86	3rd	21	14	YES
17.	Lack of complement any school based club or organization in agricultural science	139	91.3	2nd	13	8.7	YES
18.	Continuous assessment not based on school farm work	125	83,3	5th	25	16.7	YES
19.	Examination bodies and teachers not examining skill acquired on school farms.	140	93.3	1st	10	6.7	YES

Result presented in Table 2 reveals that respondents agreed with 14 out of the 19 items (91% to 53%) they disagree with 5 items (09% to 46%). The respondents agreed that the major cause of decline is that examination bodies such as WAEC, NECO, NABTEB and teachers examine skills acquired on school farms (93.3%) lack of complementary school-based club or organization in agricultural science (91%). Practical agriculture not made interesting (86%). They most disagreed that shortage in trained staff to teach practical agriculture.

Research Question 2: What are the measures to improve the use and operation of school farm for skill development in agriculture in secondary schools?

Table 3: Measures to improve the use and operation of school farm for skill development.

S/N	Measures to improve the utilization and operation of school farm for skill development	YES		NO		REMARKS
		N	%	N	%	

		1	2	3	4	
1.	Only trained agricultural teachers should teach agricultural science in secondary schools.	145	96	05	4	YES
2.	Teachers on ground should be retrained.	125	83	25	17	YES
3.	Improvement on government policy to favour practical agriculture.	127	85	23	15	YES
4.	SSCE final examinations should examine skills acquired on the farm.	137	91	13	9	YES
5.	More fund and funding should be provided/made available to finance school farm work.	93	62	47	38	YES
6.	Involvement of agencies like ADP, Board for Vocational and Technical Education and Ministry of Education to evaluate school farm work.	99	66	41	34	YES
7.	Flexibility of time for agric use of school facilities and personnels on the part of school administration.	102	68	48	32	YES
8.	Adjustment of school time table for special farm operations.	77	51	73	49	YES
9.	Teacher should introduce innovations and new technologies to make farm works easier.	89	59	61	41	YES
10.	Government/PTA philanthropics should supply farm tools and implements to schools.	130	86	20	14	YES
11.	Present examination system should change from theory oriented to practical-field-projects.	147	98	03	2	YES
12.	Practicals on students crops and livestock pens management to form a percentage of the Agric Examination continuous assessment.	143	95	07	5	YES
13.	School farm work should emphasis keeping of daily and routine farm records.	138	92	12	8	YES
14.	Adoption of 4years crop rotation principle to manage the farm.	141	93	10	7	YES

Data in table 3 shows that the most important measures to improve the use of school farm is for the examination bodies and teachers to change from theory oriented examinations to practical field ones (98%). All the measures have a percentage 'Yes' of between 98% and 51%, showing that all the items were accepted as measures that can be used to improve the utilization and operation of school farm for skill development in secondary schools in Ekiti State.

Research Question 3: What are the trends in the use of school farm?
Table 4: Trends in the use of school farm.

Teacher uses school farm to complement teaching agric in	Always	Often	Sometimes	Never	Total
Public schools	11 (11%)	16 (16%)	22 (22%)	51 (51%)	100
Private schools	07 (14%)	10 (20%)	11 (22%)	22 (44%)	50

Results presented in Table 4, shows that (51% and 44%) of Agric teachers in public and private secondary school never use the school farm to complement the teaching of Agric in secondary schools in Ekiti. Only (11% and 14%) of the teachers in public and private schools always use the school farm.

Hypothesis I: There is no significant difference in the perception of respondents (in public and private secondary schools) in the declining nature of school farm usage for skill development in Ekiti State, Nigeria.

Table 5: To Mean, standard deviation and t-value on the pattern of decline in the use of school farm for skill development in Ekiti State.

Agric Teachers in	N	Mean X	SD	df	Calculated t-value	Critical t-value	Remark
Public schools	100	29.67	6.66				
Private schools	50	28.42	7.31		1.25	1.96	Not significant

P > 0.05, df = 148, Ns Not significant

Table 5 reveals a calculated t value of 1.25 while the critical t value is 1.96. Given the calculated t-value that is less than the critical t value at 0.05 alpha level, means there is no significant difference in the perception of public and private teachers about the declining nature, the use and operation of school farm in secondary schools in Ekiti State.

III. DISCUSSION OF FINDINGS

The study found out that the major causes of decline in the use of school farm for skill development include the fact that class teachers and examination bodies like WAEC, NECO, NABTEB do not examine competencies acquired or skills acquired on school farm, and lack of school based club or organization in agriculture in secondary schools. This finding is supported by Olaitan and Mama 2001 that most students who pass external examinations such as school certificate examinations, WAEC or NECO in agricultural science do so without practicing on the farm. The students record success at internal and external examinations in Agricultural Science by studying agricultural theoretically without the technical know how or practical orientation through the farm. Similarly, Okorie and Famiwole (1997) found out that in-school youth organizations in agriculture should be encouraged and established in all secondary schools (NCCE, 2008), because its activities complements the classroom instructions in agriculture, and students are trained in the organization to put into practice what they have learn in the classroom through a guided 'learning by doing' approach to education.

The study also revealed that the both teachers and the examination bodies should change from theory oriented examinations to practical field work. One of the theory of vocational education, and a principle of school farm management as profounded by Olaitan and Mama is that, the school farm should train students directly and specifically in the thinking habits, and manipulative skills required on farm production/operations outside this school. Changing from theory oriented examinations to practical field work on the farm will develop students skills and make them more competent to face the challenges require to establish a farm or gardening on their own in future. According to them specific learning experience in the school farm will enable students to form the right habits and thinking necessary for success in any relevant agricultural occupations later in life. The findings of the study also revealed that most Agric teachers in both private and public secondary school in Ekiti randomly or never use or operate the school farm for teaching and learning Agricultural Science. This findings was also collaborated by the findings of Olaitan and Mama 2001 that in most schools and colleges, the school farm is not well developed and utilized to develop skills in students. This appears to be an assertion.

Professionally, the school farm and its management practices should be a replica of farms enterprise where students will work after graduation. This principles, is backed by the directive of the National Policy on Education (2004) and National Curriculum for Secondary Schools (2008) that all schools should establish a school farm with arable crops and two (2) types of livestock. The body also recommended the use of 4year crop rotation farming system to conserve the soil. In which case, all secondary schools in Ekiti State are expected to establish a standard and school farm.

The result of the hypothesis revealed a no significant difference in the perception of both the Agric teachers in public and private secondary schools about the decline in the use and operation of school farm in secondary schools in the state. This findings agree with the earlier observations of Olaitan and Mama (2001) that the teachers of agriculture are the vehicle for successful agricultural development in the schools. The

teachers are the managers of the school farm; therefore they are very important and relevant to the economic development of the agrarian society and skill development in agric students.

IV.   CONCLUSION

The school farm is the main agricultural laboratory, specifically designed for the purpose of imparting knowledge, skill, and farm habits to students studying agriculture in schools and colleges through learning by doings. This study confirms that only few secondary schools have well developed school farm or garden for skill development purpose. Most of the schools never have or some seldom use the school farm to complement classroom instruction. Many causes had been identified in this study, few of the most important caused of the decline in the use and operation of standard school farm are; that teachers and examination bodies such as WAEC, NECO, NABTEB do not examine the skill acquired on school farms; lack of school based agricultural clubs or organization and practical agricultural science are not made interesting. The study has established the need to improve the declining nature of the use of school farm.

V.   RECOMMENDATIONS

Based on the findings of this study, the following recommendations are made.

[1]   All secondary schools offering agricultural science should be compelled to establish and operate standard school farms.

[2]   The Ministry of Education in collaboration with other relevant agencies such as ADP, BVTE, Agric input & supplies should give specifications for a standard school farm and carryout periodic evaluation of school farm work every term, planting season and academic session.

[3]   Practical agriculture should be made interesting to students. Students should be entitled to some fraction of the proceed from their plot/work on the school farm.

[4]   A school-based youth organization in agriculture should be established and operated in all secondary schools in the state. The activities of the organization should complement classroom instruction in agriculture through learning by doing and experiential activities.

[5]   The training programmes and conferences should be organized for all agric science/education teachers presently on the field to upgrade them and expose them to modern, innovative and integrated approach to school farm management.

[6]   The use of 4 year crop rotational system should be emphasized while operating the school farms in all secondary schools.

[7]   The PTA, philanthropist and successful agro-based businessmen should contribute to school farm development.

[8]   Time tables in secondary schools should be made to favour the practice of agriculture because of timeliness of operation.

[9]   The state, local and federal government should fund school farms and supply relevant simple tools and implements to make working on the farm easier.

[10]   All students of agriculture should be encouraged to keep farm record books.

REFERENCES

[1]   Agriculture and rural development ISSN 1619-8891 www.ruraldevelopment.de.
[2]   Famiwole, R.O. and Okorie, J.U. (1997). *Developing Policy and Criteria for Establishing a Viable In-School Youth Organization in Agriculture*. Ph.D Thesis, Faculty of Education, University of Nigeria, Nsukka.
[3]   National Curriculum for Secondary Schools (2008) Federal Ministry of Education Abuja: Nigeria.
[4]   National Education Research Development Council NERDC (2008). Federal Ministry of Education Abuja: Government Press.
[5]   National Policy on Education (2004) Federal Ministry of Education Abuja: Government Press.
[6]   Olaitan, S.O. (1984). *Agricultural Education in the Topics*: Basing-Stoke. London, Macmillan Publishers.
[7]   Olaitan, S.O. (1984). *The Image of the Teacher of Agriculture on Secondary Schools in Ondo State*. Unpublished Research Report. Nsukka: University of Nigeria.
[8]   Olaitan, S.O. (1996). *Vocational and Technical Education in Nigeria. Issues and Analysis*. Onitsha: Cape Publisher.
[9]   Olaitan, S.O. and Mama, R.O. (2001). *Principles and Practice of School Farm Management*. Owerri: Cape Publishers.
[10]   Seweje, R.O. and Jegede, S.A. (2005). *Science Education and Science Teaching Methods*. Lagos: Atlantic Associated Publishers.
[11]   Teaching Service Commission, (2012). Ekiti State Ministry of Education, Ado-Ekiti.

Anomaly Detection Using Hidden Markov Model

[1,]Sonali N.Jadhav, [2,]Kiran Bhandari

[1]M.E Scholar, [1,2],Assitant Professor, Thakur College Of Engineering And Technology,
Mumbai, India

ABSTRACT:

This paper proposes an empirical method of Anomaly detection by analyzing the spending habit of vendee. Proposed system models the sequence of operations in credit card transaction processing using a Hidden Markov Model (HMM) and shows how it can be used for the detection of frauds. In the existing credit card fraud detection business processing system, fraudulent transaction will be detected after transaction is done. It is difficult to find out fraudulent and regarding loses will be barred by issuing authorities. Hidden Markov Model is the statistical tools for engineer and scientists to solve various problems. It is shown that credit card fraud can be detected using Hidden Markov Model during transactions. Hidden Markov Model helps to obtain a high fraud coverage combined with a low false alarm rate. Does not require fraud signatures and yet is able to detect frauds by considering a cardholder's spending habit. Card transaction processing sequence by the stochastic process of an HMM. The details of items purchased in Individual transactions are usually not known to an FDS running at the bank that issues credit cards to the cardholders. Hence HMM is an ideal choice for addressing this problem. The objective of the system is to detect the Anomaly during the transaction only and confirm the fraud by asking some security code. Hidden Markov Model helps to obtain a high fraud coverage combined with a low false alarm rate.

KEYWORDS: *Anomaly detection, Hidden Markov Model, Online Transactions, Data Mining, Clustering,Probability,Security.*

I. INTRODUCTION

In day to day life credit cards are used for purchasing goods and services with the help of virtual card for online transaction or physical card for offline transaction. In physical-card based purchase, the cardholder presents his card physically to a merchant for making a payment. To carryout fraudulent transactions in this kind of purchase, an attacker has to steal the credit card. If the cardholder does not realize the loss of card, it can lead to a substantial financial loss to the credit card company. In online payment mode, attackers need only little information for doing fraudulent transaction (secure code, card number, expiration date etc.).In this purchase method, mainly transactions will be done through Internet or telephone. To commit fraud in these types of purchases, a fraudster simply needs to know the card details. Most of the time, the genuine cardholder is not aware that someone else has seen or stolen his card information. The only way to detect this kind of fraud is to analyze the spending patterns on every card and to figure out any inconsistency with respect to the "usual" spending patterns. Fraud detection based on the analysis of existing purchase data of cardholder is a promising way to reduce the rate of successful credit card frauds [1]. Since humans tend to exhibit specific behaviorist profiles, every cardholder can be represented by a set of patterns containing information about the typical purchase category, the time since the last purchase, the amount of money spent, etc. Deviation from such patterns is considered as fraud.

A Hidden Markov Model [5] is a finite set of states; each state is linked with a probability distribution. Transitions among these states are governed by a set of probabilities called transition probabilities. In a particular state a possible outcome or observation can be generated which is associated symbol of observation of probability distribution. It is only the outcome, not the state that is visible to an external observer and therefore states are ``hidden" to the outside; hence the name Hidden Markov Model. Hence, Hidden Markov Model is a perfect solution for addressing detection of fraud transaction through credit card. One more important benefit of the HMM-based approach is an extreme decrease in the number of False Positives transactions recognized as malicious by a fraud detection system even though they are really genuine.

In this prediction process, HMM consider mainly three price value ranges such as.1) Low (l),2) Medium (m) and,3) High (h).First, it will be required to find out transaction amount belongs to a particular category either it will be in low, medium, or high ranges.Initially the HMM is trained with the normal behavior of a card holder then spending patterns of user can be determined with the help of K-means clustering algorithm. If an incoming transaction is not accepted by the HMM with sufficient probability then it can be detected as fraud for further confirmation security question module will be activated that contains some personal questions that are only known to authorized customer and if the transaction is fraudulent then verification code is asked for further confirmation. Hidden Markov model works on Markov chain property in which probability of each subsequent state depends on the previous state, which consists of observation probabilities, transition probabilities and initial probabilities. A hidden Markov model can be considered a generalization of a mixture model where the hidden variables (or latent variables), which control the mixture component to be selected for each observation, are related through a Marko process rather than independent of each other.

II.   RELATED WORK

Ghosh and Reilly [3] have proposed credit card fraud detection with a neural network. They have built detection system, which is trained on a large sample of labeled credit card account transactions. These transactions contain example fraud cases due to lost cards, stolen cards, application fraud, counterfeit fraud, mail-order fraud, andnonreceived issue (NRI) fraud.Stolfo et al [4] suggest a credit card fraud detection system (FDS)using meta-learning techniques to learn models of fraudulent credit card transactions. Metalearningis a general strategy that provides a means for combining and integrating a number of separately built classifiers or models. A meta-classifier is thus trained on the correlation of the predictions of the base classifiers. The same group has also worked on a cost-based model for fraud and intrusion detection [11]. They use Java agents for Meta-learning (JAM), which is a distributed data mining system for credit card fraud detection. A number of important performance metrics like TP-FP (True Positive – False Positive) spread and accuracy have been defined by them.Aleskerov et al. [12] present CARDWATCH, database mining system used for credit card fraud detection. The system, based on a neural learning module, provides an interface to a variety of commercial databases. Fan et al. [13] suggest the application of distributed data mining in credit card fraud detection. Braise et al. Stolfo. use an agent-based approach with distributed learning for detecting frauds in credit card transactions. It is based on artificial intelligence and combines inductive learning algorithms and metal earning methods for achieving higher accuracy. Phua et al. [14] suggest the use of met classifier similar to in fraud detection problems.

They consider naive Bayesian, and Back Propagation neural networks as the base classifiers. A met classifier is used to determine which classifier should be considered based on skewness of data. Although they do not directly use credit card fraud detection as the target application, their approach is quite generic.Vatsa et al.[15] have recently proposed a game-theoretic approach to credit card fraud detection. They model the interaction between an attacker and an FDS as a multi stage game between two players, each trying to maximize his HMM-based applications are common in various Areas such as speech recognition, bioinformatics, and Genomics. In recent years, Joshi and Phoba [16] have investigated the capabilities of HMM in anomaly detection. They classify TCP network traffic as an attack or normal using HMM. Cho and Park [17] suggest an HMM-based intrusion detection system that improves the modeling time and performance by considering only the privilege transition flows based on the domain knowledge of attacks. Ours ton etal. [18] have proposed the application of HMM in detecting multistage network attacks. Hoang et al. present a new method to process sequences of system calls for anomaly detection using HMM. The key idea is to build a multilayer model of program behaviors based on both HMMs and enumerating methods for anomaly detection. Lane has used HMM to model human behavior. Once human behaviorism correctly modeled, any detected deviation is a cause for concern since an attacker is not expected to have a behavior similar to the genuine user. Hence, an alarm is raised in case of any deviation.

III.   PROPOSED WORK

3.1 Problem Definition and Outline

In proposed system, by using Hidden Markov Model (HMM) which does not require fraud signatures and yet is able to detect frauds by considering a cardholder's spending habit. Card transaction processing sequence by the stochastic process of an HMM. The details of items purchased in Individual transactions are usually not known to an FDS running at the bank that issues credit cards to the cardholders. Hence HMM is an ideal choice for addressing this problem. To complete the transaction Vendee should answer the security questions. Fraud is confirmed by asking some security code which is sent by email transaction proceed only

when verification code is correct otherwise transaction cancelled. Fraud is detected using the probability difference that is in between old observation sequence and new observation sequence.
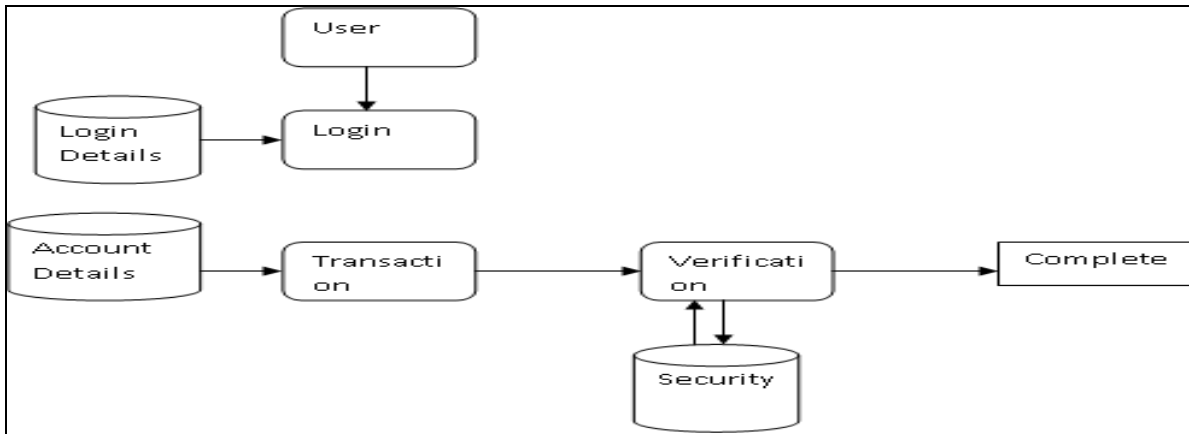
Fig.1: Outline of the proposed method

3.2. Hidden Markov Model

HMM is basically a model consisting of sequence of states that works on Markov chain property [5]. Name Hidden here indicates that observer does not know in which state it is but having a probabilistic insight on where it should be. Input to HMM is observation sequence and output is probability [6] of a sequence. A hidden Markov model can be considered a generalization of a mixture model where the hidden variables (or latent variables), which control the mixture component to be selected for each observation, are related through a Markov

process rather than independent of each other
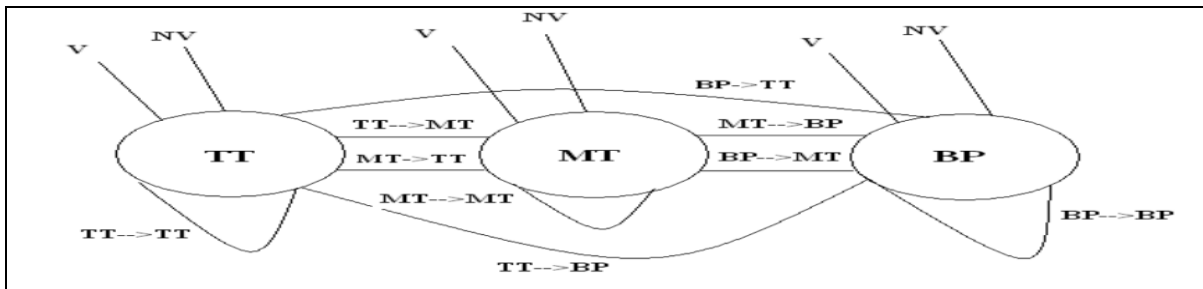
Fig 2: HMM model for fraud detection

Here in above Figure 2 HMM Model for fraud detection is considered [8]. Three Different kind of Purchases are shown they are represented as states of HMM TT (Travel Ticket),MT(Movie Ticket),BP(Book Purchase). V and NV are two Observation Symbols either of one will active for particular state they are shown on each state. V indicates Violation means if incoming transaction violates the Behavior sequence then V will be observed symbol to that state and OTP is sent to the Customers Mobile Number. NV indicates on-Violation means there is no anomaly and incoming Transaction is Normal. No action is performed in this Observation. All others lines and curves indicates Transition. From one state to another state. States will increase if Sequence of number of Purchase made is increased. to use HMM we need to calculate the HMM parameters such as state and transition probabilities. Those parameters are calculated using Baum-Welch algorithm [9].Baum and Welch algorithm is given below.
Baum-Welch Algorithm::
1. Initialize the parameters (state, transition) to some values
2. Calculate "forward-backward" probabilities based on the Current parameters
3. Use the forward-backward probabilities to estimate the Expected frequencies
4. Use the expected frequencies to estimate the parameters.
5. Repeat 2 to 4 until the parameters converge.

3.3. K-Means Clustering Algorithm

By using K-MEANS clustering algorithm which divides the spending profile of a user into low medium and high cluster and accordingly generates observation symbols that are further given to HMM for training as well as detection purpose K-means clustering algorithm first divides the transaction amount into different clusters. Consider example 10 transactions,)

Tabel 1.Transaction amount

Transaction	1	2	3	4	5	6	7	8	9	10
Amount	40	25	15	5	10	25	15	20	10	80

Table 2. Output of K-means clustering

Cluster mean/centroid name	Cl	Cm	Ch
Observation symbol	V1=l	V2=m	V3=h
Mean value	8.3	20	60
Percentage of total transactions	30	50	20

By applying K-Means clustering algorithm on above transactions output is shown in table 2.Considering M= 3, if we execute K-means algorithm on the example transactions in Table 3.2, we get the clusters, as shown in Table 3.3, with cl, cm, and ch as the respective centroids. It may be noted that the dollar amounts 5, 10, and 10 have been clustered together as cl resulting in a centroid of 8.3. The percentage (p) of total number of transactions in this cluster is thus 30 percent. Similarly, dollar amounts 15, 15, 20, 25, and 25 have been grouped in the cluster cm with centroid 20, whereas amounts 40 and 80 have been grouped together in cluster ch. cm and ch, thus, contain 50 percent and 20 percent of the total number of transactions. When the FDS receives a transaction T for this cardholder, it measures the distance of the purchase amount x with respect to the means cl, cm, and ch to decide the cluster to which T belongs and, hence, the corresponding observation symbol. As an example, if x= $10, then in Table 2 the observation symbol is V1= l.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The experimental results include identifying the spending habit of customer and detect the fraud by analyzingthe spending habit of customer and then precede the transaction. Implementation is done using java and MySQL. Some of the snapshots for the implementation given below:

Step1: Vendee will select the product from list and add it to cart.initialy cardholder will select the product and add it to the cart.
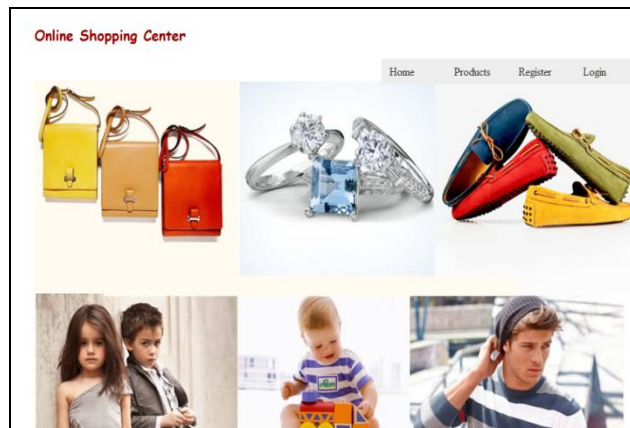
Fig 3: Selection of product

Fig 4: Purchase the product

Step2: Enter Details after selecting the product.

Fig 4: Card Details

Fig 5: States of HMM

As shown in figure 4 vendee will enter the Details when he submit the details using k-means clustering algorithm observation sequence is generated Low Medium High further Hidden Markov Model starts calculating the spending habit of customer figure 5 shows states of Hidden Markov Model.

Step3: comparing previous observation sequence with existing and calculates the probability difference if it is >0 then fraud is detected and mail sent to Vendee for confirmation shown in figure 7, else the transaction is completed.

Fig 6: Probability sequence generated by HMM

Fig 7: Anomaly Detection by HMM

Step4: Fraud detection confirmed by asking verification code if user will enter the correct verification code then transaction is completed else fraud is confirmed.
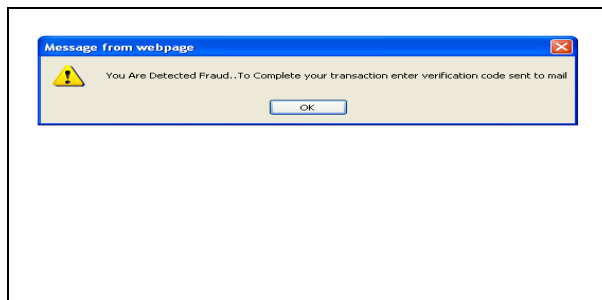
Fig 8: Fraud detection and confirmation.

Table 3. List of all the most recent transaction

Transaction	1	2	3	4	5	6	7	8	9	10
Amount	140	125	15	5	10	125	15	120	10	280
Transaction	11	12	13	14	15	16	17	18	19	20
Amount	210	550	800	110	35	118	20	148	141	6

In Table 3, list of all the most recent transaction is placed at the first position and correspondingly first transaction is placed at the last position in the table. The pattern of spending profile of the card holder is shown in Figure 9 based on all transactions done. Transactions are shown.
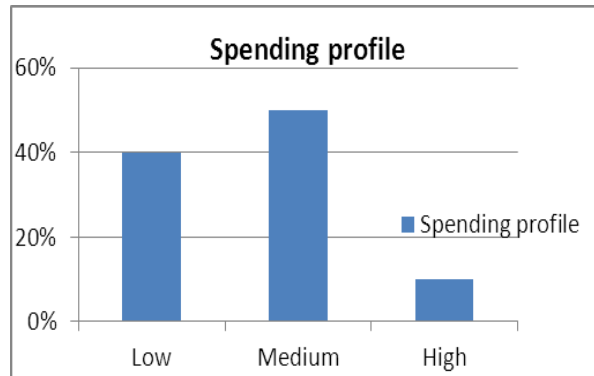
Fig. 9: Spending profile of all transaction
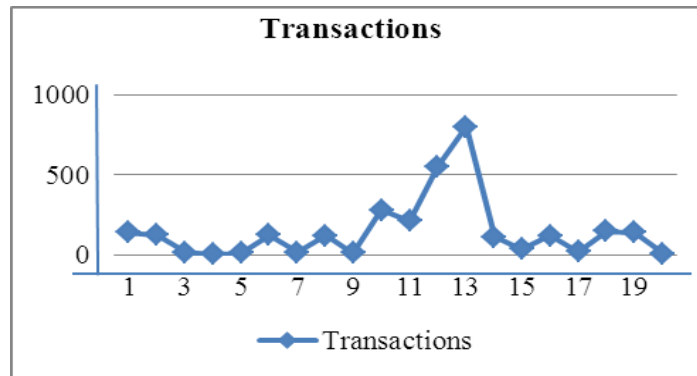
fig.10 Percentage of each spending profile
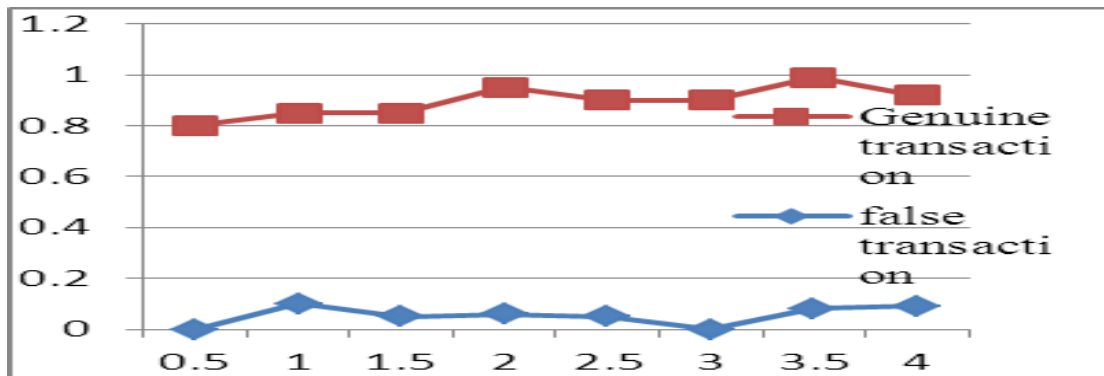
Fig:11 Probability of False Alarm compared with Fraud Transaction Mean Distribution

Fraud detection mean distribution is shown in Figure 11, where probability of false transaction compared with that of genuine transaction. Hence, it is expected that when the probability of false alarm will be more than threshold probability, then it should generate an alarm for fraudulent and also decline the transaction.

V. CONCLUSION.

We proposed system which is an application of HMM in Anomaly Detection. The different steps in credit card transaction processing are represented as the underlying stochastic process of an HMM. The ranges of transaction amount can be used as the observation symbols, whereas the types of item have been considered to be states of the HMM. Also proposed system suggests a method for finding the spending profile of cardholders, as well as application of this knowledge in deciding the value of observation symbols and initial estimate of the model parameters. It has also been explained how them can detect whether an incoming transaction is fraudulent or not. The system is also scalable for handling large volumes of transactions. The proposed method can be enhanced to achieve more accuracy and better algorithms for clustering.

REFERENCES

[1]   "Study on Fraud Risk Prevention of Online Banks" By Qinghua Zhang. 2010 International Conference onNetworks Security, Wireless Communications and   Trusted Computing.

[2]   "Fraudulent Internet Banking Payments Prevention using Dynamic Key" By Osama Dandash Yiling Wang andPhu Dung Leand Bala Srinivasan. "JOURNAL OF
NETWORKS, VOL. 3, NO. 1, JANUARY 2008".

[3]   Ghosh, S., and Reilly, D.L., 1994. Credit Card Fraud Detection with aNeural-Network, 27th Hawaii International l Conference on InformationSystems, vol. 3 (2003), pp. 621- 630.

[4]   Stolfo, S. J., Fan, D. W., Lee, W., Prodromidis, A., and Chan, P. K., 2000.Cost-Based Modeling for Fraud and Intrusion Detection: Results from theJAM Project, Proceedings of DARPA Information SurvivabilityConference and Exposition, vol. 2 (2000), pp. 130-144.

[5]   L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in SpeechRecognition," Proceedings of the IEEE, vol. 77, no. 2, pp. 257-286, 1989.

[6]   K. S. Trivedi, Probability and Statistics with Reliability, Queuing, and Computer Science Applications, Second Edition, John Wiley and Sons, New York, 2001.

[7]   Iyer, Divya.; Mohanpurkar, Arti; Janardhan, Sneha; Rathod, Dhanashree; Sardeshmukh, Amruta," Credit Card Fraud Detection" Proceedings of the IEEE, pp. 1062-1066, 2011.

[8]   Jiawei Han and Michelin Kamber,"Data mining concepts and Techniques",Elsevler publication second edition,pp 598.

[9]   "Hidden Markov Model and Baum-Welch algorithm" by Lloyd R. Welch IEEE Information Theory Society Newsletter Vol. 53, No.4, December 2003.

[10]   S.B. Cho and H.J. Park, "Efficient Anomaly Detection by Modeling Privilege Flows using Hidden Markov Model," Computer and Security, vol. 22, no. 1, pp. 45-55, 2003.

[11]   R. Brause, T. Langsdorf, and M. Hepp, "Neural Data Mining for CreditCard Fraud Detection," roc. IEEE Int'l Conf. Tools with ArtificialIntelligence, pp. 103-106, 1999.

[12]   Aleskerov, E., Freisleben, B., and Rao, B., 1997. CARDWATCH: ANeural Network Based Database Mining System for Credit Card FraudDetection, Proceedings of IEEE/IAFE: Computational Intelligence forFinancial Eng. (1997), pp. 220-226.

[13]   W. Fan, A.L. Prodromidis, and S.J. Stolfo, "Distributed Data Mining inCredit Card Fraud Detection," IEEE Intelligent Systems, vol. 14, no. 6,pp. 67-74, 1999.

[14]   C. Phua, D. Alahakoon, and V. Lee, "Minority Report in FraudDetection: Classification of Skewed Data," ACM SIGKDD ExplorationsNewsletter, vol. 6, no. 1, pp. 50-59, 2004.

[15]   V. Vatsa, S. Sural, and A.K. Majumdar, "A Game-eoretic Approach toCredit Card Fraud Detection," Proc. First Int'l Conf. Information SystemsSecurity, pp. 263-276, 2005

[16]   S.S. Joshi and V.V. Phoha, "Investigating Hidden Markov ModelsCapabilities in Anomaly Detection," Proc. 43rd ACM Ann. SoutheastRegional Conf., vol. 1, pp. 98-103, 2005.

[17]   S.B. Cho and H.J. Park, "Efficient Anomaly Detection by ModelingPrivilege Flows Using Hidden Markov Model," Computer and Security,vol. 22, no. 1, pp. 45-55, 2003.

[18]   D. Ourston, S. Matzner, W. Stump, and B. Hopkins, "Applications ofHidden Markov Models to Detecting Multi-Stage Network Attacks,"Proc. 36th Ann. Hawaii Int'l Conf. System Sciences, vol. 9, pp. 334-344,2003.

Survey Of Currency Recognition System Using Image Processing

Amol A. Shirsath[1] S. D. Bharkad[2]

[1]*PG Student, Government College of Engineering, Aurangabad (M.S.),*
[2]*Assistant Professor, Dept. of Electronics & Tele-communication, Government College of Engineering,*

ABSTRACT

Image processing based currency recognition technique consists of few basic steps like image acquisition, its pre-processing and finally recognition of the currency. Normally camera or scanner is used for image acquisition. Then these images are processed by using various techniques of image processing and various features are extracted from the images which are the key concept behind currency classification. Application area of currency recognition includes foreign exchange, automatic selling of things and in banks. Recognition ability depends on the currency note characteristics of particular country and extraction of features.

KEYWORDS : *Classifier, Feature extraction, Foreign exchange, Markov chain , LBP, Neural Network, Pre-processing,*

I.   INTRODUCTION

In recent years, along with the accelerative developments of world economics incorporation course, the start of euro area, and the increase of asia economics, frontier trade and personal intercourse of various countries are frequent increasingly. Travelling people always take many countries of paper currency. Probabilities that the paper currencies of various countries are probably interweaved together therefore rises increasingly. It is a challenge for conventional paper currency recognition systems.

There are approximately 50 currencies all over the world, with each of them looking totally different. For instance the size of the paper is different, the same as the colour and pattern. The staffs who work for the money exchanging (e.g. Forex Bank) have to distinguish different types of currencies and that is not an easy job. They have to remember the symbol of each currency. This may cause some problems (e.g. wrong recognition), so they need an efficient and exact system to help their work. Also paper currency became older than coins and also possibility of joining broken currency is greater than that of coin currency. So the aim of Currency recognition system is to help people who need to recognize different currencies, and work with convenience and efficiency.Currency recognition is an image processing technology that is used to identify currency of various countries.

Probabilities that the paper currencies of various countries are probably interweaved together therefore rises increasingly. It is a challenge for conventional paper currency recognition systems. However, the focus of most of the conventional currency recognition systems and machines is on recognizing counterfeit currencies. It is not enough for practical businesses. The reason is that in most of banks, especially those internationalized banks, there are large quantities of cash belonging to many different countries need to be process, and it is possible that all of them are real cashes. The situation that cashes belonging to different countries mixes together is possible to occur. It cannot be processed with conventional currency recognition systems. Paper currency recognition systems should be able to recognize banknotes from each side and each direction. Since banknotes may be defaced during circulation, the designed system should have a meaningful accuracy in detecting torn or worn banknotes. The technology of currency recognition is used to research the visible and hidden currency characters, identify features all-around and dispose of the process on time. The original information has a loss because paper currency will get to the worm and blurry, even damaged by human being in circulation.

II.   SYSTEM ARCHITECTURE OF PAPER CURRENCY RECOGNITION

System architecure of paper curency recognition includes five stages Image acquisition, pre-processing, feature extraction, classification, and result. Let us explain each of these stages one by one.

A. Image Acquisition

There are various ways to acquire image such as with the help of camera or scanner. Acquired image should retain all the features. You can choose between two main types of cameras – analog and digital. Digital cameras can be further classified into parallel digital, Camera Link, and IEEE 1394.Scanner is a device that optically scans images, printed text, handwriting, or an object, and converts it to a digital image. Scanners typically read red-green-blue colour (RGB) data from the array. This data is then processed with some proprietary algorithm to correct for different exposure conditions and sent to the computer via the device's input/output interface.
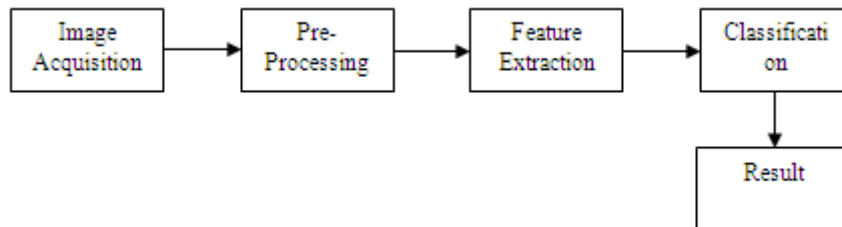
Figure 1 System Architecture of Paper Currency Recognition

B. Pre-Processing

The aim of image pre-processing is to suppress undesired distortions or enhance some image features that are important for further processing or analysis. It includes

Image Adjusting

When we get the image from a scanner, the size of the image is so big. In order to reduce the calculation, we decrease the size of image. Image Adjusting is done with the help of image interpolation. Interpolation is the technique mostly used for tasks such as zooming, rotating, shrinking, and for geometric corrections.There are two types of interpolation Bilinear and Bicubic. In the first one concept of four nearest is used to estimate the intensity at a given location. Let (x,y) denotes coordinates of the location where we want to assign an intensity value and Z(x,y) denote that intensity value so to evaluate assigned value we use equation

$$Z(x,y) = ax + by + cxy + d \qquad (1)$$

Where four coefficients can be obtained from the four equations from the four unknown can be written using four nearest neighbors of point.

Image smoothning

When using a digital camera or a scanner and perform image transfers, some noise will appear on the image. Image noise is the random variation of brightness in images. Removing the noise is an important step when image processing is being performed. However noise may affect segmentation and pattern matching.When performing smoothing process on a pixel, the neighbour of the pixel is used to do some transforming. After that a new value of the pixel is created. The neighbour of the pixel is consisting with some other pixels and they build up a matrix, the size of the matrix is odd number, the target pixel is located on the middle of the matrix. Convolution is used to perform image smoothing. Also image smoothning can be done with the help of median filter which more effective than convolution when goal is to simultaneously reduce the noise preserving edgees. Median filter replaces a pixel via the median pixel of all the neighborhoods:

$$Y[m,n] = median\{x(i,j) \ , (i,j) \in W\} \qquad (2)$$

Where W represents a neighbouhood centered around the location

III.   FEATURE EXTRACTION

It is the most challenging work in digital image processing. During the feature extraction process the dimensionality of data is reduced. Extracting too many features will not only increase the cost but also sometimes lower the system performance in terms of execution time.Feature extraction is a important procedure considerably for currency recognition, which effects on design and performance of the classifier intensively [1]. If the selected features are so large, it can easily construct a classifier with good recognition performance. The main task here is how to find effective feature among many features.

There are mainly two types of features [1] as, Structural feature: It describes geometrical and topological characteristics of pattern by representing its global and local properties. Statistical Features: It describes characteristic measurements of the pattern.In [2] component-based framework for banknote recognition by using Speeded Up Robust Features (SURF) is used. The component-based framework is effective in collecting more class-specific information and robust in dealing with partial occlusion and viewpoint changes. Furthermore, the evaluation of SURF demonstrates its effectiveness in handling background noise, image rotation, scale, and illumination changes.Size, Colour, text these are the three characteristics of a currency used to distinguish between different banknote denominations.

A. Texture Feature
For texture feature LBP (Local Binary Pattern) operator [3] is used. Texture is the visible feature of the paper currency.

LBP operator: It is originally introduced by Ojala et al [3]. In LBP, the neighbourhood pixels are converted to binary code 0 or 1 by using the gray value of the centre pixel as threshold and further arranged to form as a ordered pattern. The feature extracted with LBP gives the relationship of the texture within local area. LBP code for pixel p is defined as

$$LBP(p) = \sum_{i=0}^{7} 2^i s(g_i - g_p) \quad (3)$$

Where
g_p – Gray value of the centre pixel p.
g_i – Gray value of the ith pixel, 8-neighbourhood of p
i – 0, 1 ...7.
$s(t) = g_i - g_p$, is the threshold function and given by

$$s(t) = \begin{cases} 1, t \geq 0 \\ 0, else \end{cases} \quad (4)$$

From equation (1) we can say that LBP can produce 256 kinds of different outputs, corresponding to 256 kinds of different binary patterns.
Junfang Guo, Yanyun Zhao[4] has proposed the improved method for texture analysis using LBP. They have segment the whole image into M * N blocks. In each block calculate the LBP value for every pixel, and make the histogram of the block, which is known as a block histogram. The block histogram is normalized by the number of pixels in the block.

Markov Chain Concept: As in many countries colour spectrum and size of some banknotes are very close to each other. For such a type of difficulties we are considering template of the banknotes. And to recognize these templates we are using Markov Chain Concept [5] to represent the random phenomenon.
A random process $\{x_k, k = 0, 1, 2....\}$ is called Markov chain if the possibility value in state x_{n+1} depends on the possible value in state x_n, given as below

$$P(x_{n+1} = \beta | x_n = \alpha, x_{n-1} = \alpha_{n-1}, ... x_0 = \alpha_0)$$
$$= P(x_{n+1} = \beta | x_n = \alpha) \quad (4)$$

This possibility can be shown by P_{ij}. The state space of a Markov chain can be shown in matrix as below

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \vdots & & \cdots & \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{bmatrix} \quad (5)$$

Where n, is the number of states in the chain. In discrete time Markov chain, the possibility value of different states in the matrix is computed as

$$P_{ij} = \frac{n_{ij}}{\sum_{k=1}^{n} n_{ik}} \quad (6)$$

Where n_{ij}, is the number of transitions from state i to state j. considering (5), matrix P can be multiplied by the denominator of (5). To obtain

$$P = \begin{bmatrix} N_{11} & N_{12} & \cdots & N_{1n} \\ N_{21} & N_{22} & \cdots & N_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ N_{n1} & N_{n2} & \cdots & N_{nn} \end{bmatrix} \quad (7)$$

This is matrix is used to differentiate between textures in different denominations [3].

B. Colour Feature

If the primary image is in RGB format, then after resizing it is converted to HSV colour space[6]. Advantage of HSV colour space is that it is closer to human conceptual understanding of colours and has ability to separate chromatic and achromatic Components. Feature extraction of a colour image can be done by analysing its colour histogram, hue, saturation, intensity (or value). In HSV (Hue, Saturation, Value) space hue distinguishes colour, Saturation is the percentage of white light added to a pure colour and value represents perceived light intensity.

C. Size

The first step of recognition of algorithm proposed by H. Hassanpour [8], considers size of the banknote. It is considered because during circulation of banknote worn and torn reduces its size, also it may increased slightly by rejoining torn banknote. Therefore they proposed decision tree as follows

$$|x - x_0| < d_x \ \& \ |y - y_0| < d_y \quad (8)$$

Where x_0 and y_0 are size of the testing paper currency, x and y are size of reference paper currency. d_x, d_y shows changes in the vertical and horizontal directions.

Classifier

After getting features of currencies, it is essential to recognize the pattern of the currencies on the base of these features, which should be practised by an effective recognition system called classifier.

The input of the classifier will be the test currency images and the output of the classifier will be three parts:

[1]   The country name that using this currency;
[2]   The denomination of the input currency image;
[3]   The front or back side of the currency.

A Neural network based recognition scheme is used for Bangladeshi banknotes [9]. The scheme can efficiently be implemented in cheap hardware which may be very useful in many places. The recognition system takes scanned images of banknotes which are scanned by low cost optoelectronic sensors and then fed into a multilayer perception, trained by back propagation algorithm, for recognition.Baiqing Sun [10] proposes a kind of currency recognition system, in which a three-layer feed forward neural network is used as a classifier, They proposed it in order to improve the performance of currency recognition system. In the feed- forward neural network a kind of Gaussian function is proposed as the activation function, which is employed in all units on the hidden layer and the output layer. The characteristics of this activation function are analyzed. It is a kind of localized function, its activation approaches zero as the distance to its centre approaches infinity. Its active region is governed by its width parameter. Just relying on this property of the proposed Gaussian, the rejection capabilities of the system can be improved.Fumiaki Takeda [11] proposed a new mask having symmetrical masked area against a axis dividing long side of paper currency in equal part, they called it as a symmetrical mask. Using this mask they have obtained same value from inverse and upright image of paper currency. Firstly geometrical meaning of symmetrical mask is shown and then procedure of the mask optimization by GA (Genetic algorithm) is given. Then neural network is constructed using optimized symmetric mask to recognize currency.As explained in [12] there are numerous models are developed to recognize paper currency. Such as auto associative network, nonlinear auto associative network, radial basis function (RBF) network. Most of them uses single neural network with one hidden layer but there are certain limitation of this networks. Firstly, single NN is not sufficient to train all aspects of currency note. Secondly, decision was made completely from a NN and not from a group of NN.

IV.   CONCLUSION

This survey article presents overview of recent work on currency recognition system. Multi Country currency recognition is a challenging work as some currency notes has common features such as colour, size. To sort out this problem several authors suggest techniques such as Markov chain concepts, Local binary pattern technique. Also this paper introduces brief overview on system architecture of currency recognition system, it includes various steps starting from image acquisition through scaner or camera then processing on digital

image with the help of image processing tools and at the end result will be shown. Classifier discussed in this paper is neural network employing guassian function.

REFERENCES

[1]   Parminder Singh Reel, Gopal Krishan, Smarti Kotwal, Image Processing based Heuristic Analysis for Enhanced Currency Recognition, International Journal of Advancements in Technology, Vol 2, No 1, January 2011.

[2]   Faiz M. Hasanuzzaman, Xiaodong Yang and Yingli Tian, Robust & Effective Component based Banknote Recognition by SURF Features, NSF grant IIS-0957016.

[3]   T Ojala,M Pietikainen, and D Harwood, A comparitive study of texture measures with classification based on feature distributions, Pattern recognition, Vol.29, No. 1, 1996, pp.51-59.

[4]   Junfang Guo, Yanyun Zhao, Anni Cai, A Reliable Method for Paper Currency Recognition Based on LBP, IEEE Transactions, Proceedings of IC-NIDC2010, 978-1-4244-6853-9/10.

[5]   M. Iosifescu, Finite Markov Processees and Their Applications, Wiley, New York, NY, 1980.

[6]   G. Trupti Pathrabe, Mrs. Swapnili Karmore, A Novel Approach of Embedded System for Indian Paper Currency Recognition, International Journal of Computer Trends and Technology- May to June Isue 2011, ISSN: 2231-2803.

[7]   Rafael C. Gonzalez, Richard E. Woods, Steven L. Eddins, Digital Image Processing Using MATLAB,2012.

[8]   H. Hassanpour, A. Yaseri, G. Ardeshiri, Feature Extraction for Paper Currency Recognition, IEEE Transactions, 1-4244-0779-6/07, 2007.

[9]   Nadim Jahangir, Ahsan Raja Chowdhury, Bangladeshi Banknote Recognition By Neural Network with Axis Symmetrical Masks, IEEE Transactions, 1-4244-1551-9/07.

[10] Baiqing Sun, Research on Rejection Capabilities of Paper Currency Recognition System with the Neural Network Employing Guassian Function, Kochi University of Technology, Japan,2006.

[11] Fumiaki Takeda, Toshihiro Nishikage, and Yoshiyuki Matsumoto, Characteristics Extraction of Paper Currency using Symmetrical Masks Optimized by GA and Neuro-Recognition of Multi-National Paper Currency, IEEE Transactions, 0-7803-4859-1/98, 1998.

[12] Masato Aoba, T. Kikuchi, Y. Takefuji, Euro Banknote Recognition System using a three layer perceptron and RBF networks, IPSJ transaction on Matematical Modeling and its Applications, vol 44, 7(8), pp. 99-109, 2003.

[13] Kalyan Kumar Debnath, Sultan Uddin Ahmed, Md. Shahjahan, A Paper Currency Recognition System Using Negatively Correlated Neural Network Ensemble, Journal of Multimedia, vol. 5, No. 6, December 2010.

[14] F. Takeda and T. Nishikage, Multiple kinds of paper currency recognition using Neural Network and application for Euro currency, In Proc. IEEE International Joint Conference on Neural Networks, pp 143-147, 2000.

[15] D. A. K. S. Gunaratna, N.D. Kodikara and H. L. Premaratne, ANN Based Currency Recognition System using Compressed Gray Scale and Application for Sri Lankan Currency Notes-SLCRec, Proceedings of world academy of science, engineering and technology, vol. 35, ISSN 2070-3740, Nov 2008.

Optical Sensor System for Hemoglobin Measurement

Rajashree Doshi [1], Anagha Panditrao [2]

[1] *Department of Instrumentation and Control, Cummins College of Engineering for Women Pune University,*
India

ABSTRACT

For complete blood count Hemoglobin (Hb) is an essential parameter. This paper presents non invasive optical technique for Hb measurement. At different wavelengths absorption coefficient of blood differs this fact is used to measure the optical characteristics of blood. In this newly developed system, principle of pulse oximetry is used. Oxygenated and deoxygenated hemoglobin absorbs different amount of light at two wavelength 660nm and 940nm. Red and IR LED are used for these particular wavelengths. Transmitted light through an area of skin on finger was detected by a transimpedance amplifier photodiode. Ratio of pulsating to non pulsating component of both red and IR signal after normalization is calculated for determination of Hb. Signal acquisition by this method is totally noninvasive. The sensors assembled in this investigation are fully integrated into wearable finger clips.

KEYWORDS: absorption of light, blood, hemoglobin, infra red, LED, noninvasive, optical method.

I. INTRODUCTION

Hemoglobin (Hb) is usually measured as a part of the complete blood count from a blood sample. Hemoglobin plays important role for transporting oxygen from the lungs to the other peripheral tissue of body and exchange oxygen for carbon- dioxide and then carry carbon dioxide back to lungs where it is exchange for oxygen. Hemoglobin is made up of four protein molecules, called globulin chains; each globulin chain contains an important central structure called the heme molecule. Embedded within the heme molecule is iron that is vital in transporting oxygen and carbon dioxide in our blood. An iron contained in hemoglobin is responsible for the red color of blood. If Hemoglobin level crosses the critical limits then problem occurs such as anemia for low hemoglobin and polycythemia for high hemoglobin level. Several methods are used to measure total hemoglobin content in the blood. The most common methods utilize spectrophotometric analysis of light absorbance based on Beer-Lambert law. Other method takes advantage of the varying conductivities of blood at different concentrations of blood.

In Hemoglobincyanide method hemoglobin is chemically converted to form a cyanmethemoglobin which is having maximum absorption around 540nm wavelength. The hemoglobin concentration is determined from absorption. This technique is most broadly use in laboratory. In this invasive method to measure the Hb concentration, blood is ejected from the patient and subsequently analyzed. Apart from the discomfort of ejecting blood samples, an added disadvantage of this method is the delay between the blood collection and its analysis, which does not allow real time patient monitoring in critical situations. A noninvasive method allows pain free continuous on-line patient monitoring with minimum risk of infection and facilitates real time data monitoring allowing immediate clinical reaction to the measured data. Since the near infrared light was found to penetrate a great depth into biological tissues, near-infrared spectroscopy has been developed into a noninvasive method for biomedical sensing and clinical diagnosis. Oximetry, is well known as typical example of a near-infrared application in clinic, and can be used to noninvasive measure the oxygen saturation of human blood in-vivo [2]. The absorption of whole blood in the visible and near infrared range is dominated by the different hemoglobin derivatives and the blood plasma that consists mainly of water. It is well known that pulsating changes of blood volume in tissue can be observed by measuring the absorption of light through the blood volume. This diagnostic method is known as photoplethysmography (PPG) [3]. Aldrich et al. have reported on the ability to use NIR transmission through the fingertip at a single pseudoisosbestic wavelength (905 nm) coupled with a sonomicrometer to monitor pulsatile changes in the optical path length through the finger as well as correct for interpatient variation in finger diameter [4]. An optical method for direct measurement of Hb noninvasively was reported by Jeon et al., who used a 5-wavelength diode-emitting array, but this method requires more robust detection mechanisms [5].

In this newly developed optical sensor system two different wavelengths of light uses for the measurement of Hb concentration. This non-invasive optical measurement method is based on radiation of red and near infrared light, emitted by Light Emitting Diodes (LED) at particular wavelength of 660nm and 940nm. Transmitted light through an area of skin on finger was detected by a transimpedance amplifier photodiode. Ratio of pulsating to non pulsating component of both red and IR signal after normalization is calculated for determination of Hb. Signal acquisition by this method is totally noninvasive. The sensors assembled in this investigation are fully integrated into wearable finger clips.

II. EXPERIMENTAL METHODS

Cyanmeth is the standard method for laboratory determination of haemoglobin in blood samples. The test is usually done by dissolving 20 microliters of uncoagulated blood in 5 ml of Drabkins solution. Also different clinically used methods are Spectrophotometry, Hemoglobincyanide and conductivity based method for measurement of hemoglobin. However in these methods it is required to eject the blood sample from human body and then it is tested. It causes pain to the patient and results required are delayed. In the developed technique non contact optical sensor is developed for haemoglobin measurement.

2.1. System Overview

Basic block diagram of noninvasive hemoglobin measurement system are described in figure (1).
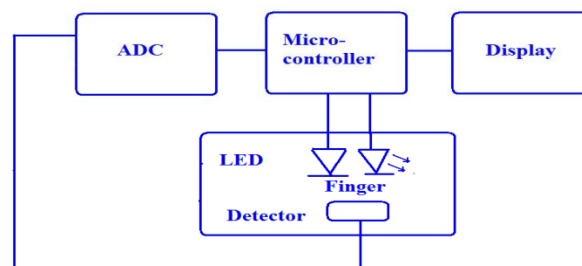
Figure1. Block diagram of hemoglobin measurement system

The non-invasive sensor systems allow a continuous measurement of the hemoglobin concentration, which is based on a pulse photometric measurement method. Thereby an area of skin on the fingertip is trans-illuminated by light which is emitted by LEDs of 660nm and 940nm. Figure (2) describe the absorption spectra for oxyhemoglobin and deoxyhemoglobin. The objective of the photometric devices described here is the non-invasive continuous measurement of heart circulation patterns and light absorbent blood components in the blood of the human finger. The arteries contain more blood during the systolic phase of the heart than during the diastolic phase, due to an increased diameter of the arteries during the systolic phase. This effect occurs only in arteries but normally not in veins. For this reason the absorbance of light in tissues with arteries increases during systole because the amount of hemoglobin (absorber) is higher and the light passes through a longer optical path length in the arteries. These intensity changes are called PPG-waves [6].

The time varying part allows the differentiation between the absorbance due to venous blood and bloodless tissue (DC part) and that due to the pulsatile component of the total absorbance (AC part). An electrical signal consisting of two components is generated by the photo detector receiving the LED emission. There is an invariant direct current (DC) component to the signal which represents ambient background light and transmission of light through invariant that is non pulsating tissues such as skin, bone and, to a certain extent, veins. The second component of the signal is an alternating current (AC) which represents varying transmission of light through the pulse varying tissues i.e. the arteries and capillaries. Both AC and DC components are affected by altered LED light intensity. Suitable wavelengths were selected for the analyses of relative hemoglobin concentration change. The principle of measurement is based on the fact of a substantial absorption/transmission difference of light in red and near infrared region between oxygenated $[HbO_2]$ and reduced hemoglobin [HHb]. HHb is optically much denser to the red light (600 ~ 750 nm) than HbO_2. whereas the reverse is true in the near infrared region (900 ~ 1000 nm) [7].

Figure2. Absorption spectra of oxy- and deoxyhemoglobin

2.2. Mathematical Implementation

Hemoglobin is a molecule in the red blood cells that has a role of delivering oxygen to tissue cells. Hemoglobin is composed of four heme groups and a protein group, known as a globin. For spectro- photometric experiments   Beer-Lambert's law is utilized and developed the notation of absorbance to express light absorption as a function of hemoglobin concentration as given in equitation:

$$OD = Log (I_0/I) = \varepsilon c L \tag{1}$$

Where OD is the optical density, I_0 is the light intensity of incident light, I *is* the light intensity of transmitted light, ε is the extinction coefficient of hemoglobin, c is the concentration of hemoglobin, and L is the length of light path through solution.

When the measured sample has a mixture of oxygenated and deoxygenated hemoglobin, equation (1) can be further expanded as,

$$OD^{\lambda} = \{\varepsilon_{HHb}^{\lambda}[HHb] + \varepsilon_{Hbo2}^{\lambda}[HbO_2]\} \, L \tag{2}$$

Where OD^{λ} is the optical density or absorbance at wavelength λ and εHHb (λ) and εHbO$_2$ (λ) are the extinction coefficients at wavelength λ for molar concentrations of deoxygenated hemoglobin, [HHb], and oxygenated hemoglobin, [HbO$_2$], respectively. By assuming light path L as 1cm. Both [HbO$_2$] and [HHb] can be determined by measuring the light absorbance at the two specific wavelengths, provided that the values for εHHb (λ) and εHbO$_2$ (λ) are known, as expressed below.

$$[HbO_2] = \frac{\varepsilon_{HHb}^{\lambda 2} OD^{\lambda 1} - \varepsilon_{HHb}^{\lambda 1} OD^{\lambda 2}}{L\left(\varepsilon_{HHb}^{\lambda 2} \varepsilon_{Hbo2}^{\lambda 1} - \varepsilon_{HHb}^{\lambda 1} \varepsilon_{Hbo2}^{\lambda 2}\right)} \tag{3}$$

$$[HHb] = \frac{\varepsilon_{Hbo2}^{\lambda 2} OD^{\lambda 1} - \varepsilon_{Hbo2}^{\lambda 1} OD^{\lambda 2}}{L(\varepsilon_{HHb}^{\lambda 1} \varepsilon_{Hbo2}^{\lambda 2} - \varepsilon_{HHb}^{\lambda 2} \varepsilon_{Hbo2}^{\lambda 1})} \tag{4}$$

It follows that changes in [HHb] and [HbO2] can be consequently given as

$$\Delta[HbO_2] = \frac{\varepsilon_{HHb}^{\lambda 2} \Delta OD^{\lambda 1} - \varepsilon_{HHb}^{\lambda 1} \Delta OD^{\lambda 2}}{L(\varepsilon_{HHb}^{\lambda 2} \varepsilon_{Hbo2}^{\lambda 1} - \varepsilon_{HHb}^{\lambda 1} \varepsilon_{Hbo2}^{\lambda 2})} \tag{5}$$

$$\Delta[HHb] = \frac{\varepsilon_{Hbo2}^{\lambda 2} \Delta OD^{\lambda 1} - \varepsilon_{Hbo2}^{\lambda 1} \Delta OD^{\lambda 2}}{L(\varepsilon_{HHb}^{\lambda 1} \varepsilon_{Hbo2}^{\lambda 2} - \varepsilon_{HHb}^{\lambda 2} \varepsilon_{Hbo2}^{\lambda 1})} \tag{6}$$

$$\Delta[Hb]total = \Delta[HHb] + \Delta[HbO_2] \tag{7}$$

Where ΔOD^{λ} represents a change in optical density at the specific wavelength, λ, and equals log (IB/IT)**.** IB and IT correspond to light intensities measured under the baseline and transient conditions [8].

III.   SENSOR DESIGN

The developed hemoglobin sensor system consist of a number of hardware modules, which include appropriate light sources, constant light intensity circuit, transimpedance amplifier, DSO microcontroller, and GUI. Figure.3 is a schematic representation of hemoglobin measurement. The sensor consist of emitter as LEDs, with centre wavelengths of $\lambda_1 = 660$nm, $\lambda_2 = 940$nm. These two wavelengths are selected because at 660nm wavelength absorbance of deoxyhemoglobin greatly exceeds the absorbance of oxyhemoglobin where as at

960nm wavelength absorbance of oxyhemoglobin greatly exceeds the absorbance of deoxyhemoglobin. These LEDs are installed in the upper shell of a finger clip.

Figure3.Schematic representation of hemoglobin measurement sensor system

Source intensity should remain constant for this constant light intensity circuit is used. To detect the transmitted light OPT101 transimpedance amplifier is used as detector. The OPT101 is a monolithic photodiode with on-chip transimpedance amplifier. This single receiver photo diode is installed in the lower shell of the finger clip. The probe is placed to the patient's body usually on the finger. Red and infrared light is then emitted sequentially through the body tissue. The transmitted light is sensed by photodiode. Out-put voltage of photodiode increases linearly with light intensity. The amplifier is designed for single or dual power-supply operation, making it ideal for battery operated equipment. Integrated combination of photodiode and transimpedance amplifier on a single chip eliminates the problems commonly encountered in discrete designs such as leakage current errors, noise pick-up, and gain peaking due to stray capacitance. The 0.09×0.09 inch photodiode is operated in the photoconductive mode for excellent linearity and low dark current. The OPT101 operates from +2.7V to +36V supplies and quiescent current is only 120μA. It is available in clear plastic 8-pin DIP, and J-formed DIP for surface mounting. Temperature range is 0℃ to + 70℃.For digitalized this analog signal 40 pin pic 16f 877A microcontroller is used which is having inbuilt ten bit analog to digital converter. The microcontroller facilities software controlled and time multiplexed operation of light sources and receiver channels.
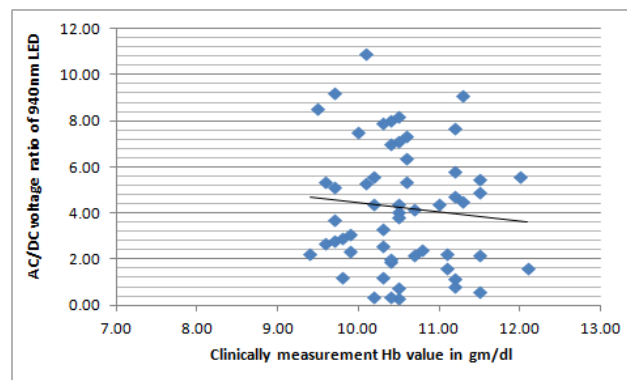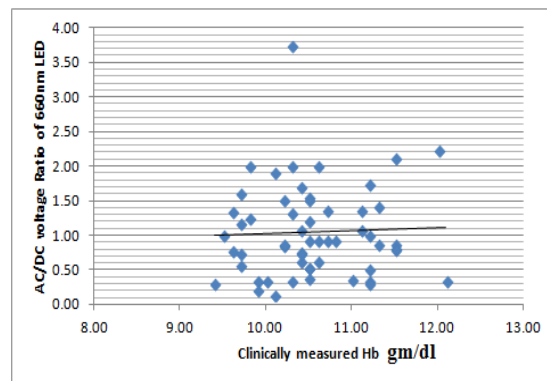
Figure4. Plot of clinically measured Hb vs. AC/DC   ratio of 660nm LED and 940nm LED

IV.   RESULTS AND DISCUSSION

An optical sensor is developed for measurement of haemoglobin by using wavelength 660nm and 940nm. Output signal are observed by sensor probe tested on various subject, and output voltage is measured also output waveform is observed on digital storage oscilloscope. A study with n=58 adult female in range of 18 to 20 year old has been performed to test the ability of this newly developed system to measure the hemoglobin content non invasive. The photometric measurements spanning 3 to 5 minutes for each subject and were store using microcontroller. Then clinically measured hemoglobin value vs. non invasively measured AC/DC ratio of 660nm wavelength are plotted as shown in figure 4. Similar graph of AC/DC ratio of 940nm and clinically measured Hb is plotted. The AC signal must be corrected for inter LED light intensity differences prior to their use for Hb calculation. The ratio of 660nm to that of 940nm is compare to clinically measured Hb value as shown in figure5.
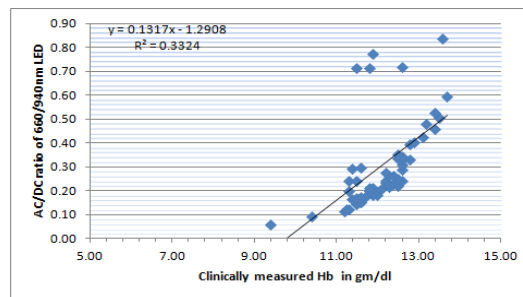
Figure5. Plot of clinically measured Hb vs.AC/DC ratio of 660nm/940nm LED

V.   Conclusion

An optical non contact type sensor for hemoglobin measurement is developed. With the help of developed technique it is possible to measure hemoglobin with two wave length 660nm and 940nm. This developed technique is tested on 60 subjects and the result shows that the ratio values increases proportionally with haemoglobin. This show the capabilities of these selected wavelengths are promising for haemoglobin measurement.

ACKNOWLEDGEMENT

REFERENCES

[1]   U. Timm, E. Lewis, D. McGrath, J. Kraitl, H. Ewald,   "LED Based Sensor System for Non-Invasive Measurement of the Hemoglobin Concentration in Human Blood", IFMBE Proceedings Vol. 23, 825-28, 2008

[2]   Suzaki, H.; Kobayashi, "Noninvasive measurement of total hemoglobin and hemoglobin derivatives using multiwavelength pulse spectrophotometry -In vitro study with a mock circulatory system" EMBS 28th Annual International Conference of the IEEE, 2006 .

[3]   J.Kraitl, H. Ewald, U.Timm "Non-invasive measurement of blood components" IEEE fifth international Conference on Sensing Technology 2011.

[4]   Aldrich TK, Moosikasuwan M, Shah SD, Deshpande KS. "Length-normalized pulse photoplethysmography:a noninvasive method to measure blood haemoglobin". Ann Biomed Eng 2002; 30:1291– 8.

[5]   Jeon KJ, Kim SJ, Park KK, Kim JW, Yoon G."Noninvasive total hemoglobin measurement".J Biomed Opt 2002;7:45–50.

[6]   Petrova, Prough, D.S.; Petrov, Brecht, "Optoacoustic technique for continuous, noninvasive measurement of total hemoglobin concentration: an in vivo study" IEMBS Volume: 1, 2004

[7]   Jae G.Kim, Mengna Xia,and Hanali Liu "Extinction coefficient of hemoglobin for near-infrared spectroscopy of tissue" IEEE Eneineering in medicine and biology magazine 2005.

[8]   Brecht, H.-P.; Petrov, "Noninvasive continuous optoacoustic monitor of total hemoglobin concentration" Engineering in Medicine and Biology Conference, Volume: 3, 2002

Dynamic Search Algorithm In Unstructured Peer-To-Peer Networks

[1,] S.Vengalakshmi M.E,   [2,]R.DhivyaB.E(M.E,)

[1,2]*Assistant Professor,Assistant Professor,Deptof Cse,Dept Of Cse, Syed Ammal Engineering College,Syedammal Engineering College,*

ABSTRACT

Designing efficient search algorithms is a key challenge in unstructured peer-to-peer networks. Flooding and random walk (RW) are two typical search algorithms. Flooding searches aggressively and covers the most nodes. However, it generates a large amount of query messages and, thus, does not scale. On the contrary, RW searches conservatively. It only generates a fixed amount of query messages at each hop but would take longer search time. We propose the dynamic search (DS) algorithm, which is a generalization of flooding and RW. DS takes advantage of various contexts under which each previous search algorithm performs well. It resembles flooding for short-term search and RW for long-term search. Moreover, DS could be further combined with knowledge-based search mechanisms to improve the search performance. We analyze the performance of DS based on some performance metrics including the success rate, search time, query hits, query messages, query efficiency, and search efficiency. Numerical results show that DS provides a good tradeoff between search performance and cost. On average, DS performs about 25 times better than flooding and 58 times better than RW in power-law graphs, and about 186 times better than flooding and 120 times better than RW in bimodal topologies.

INDEX TERMS:*Peer-to-peer, performance analysis, search algorithm.*

I.   INTRODUCTION

[I]N unstructured peer-to-peer (P2P) networks, each nodedoes not have global information about the whole topology and the location of queried resources. Because of the dynamic property of unstructured P2P networks, correctly capturing global behavior is also difficult [1], [2]. Search algorithms provide the capabilities to locate the queried resources and to route the message to the target node. Thus, the efficiency of search algorithms is critical to the performance of unstructured P2P networks [3]. Previous works about search algorithms in unstructured P2P networks can be classified into two categories: breadth first search (BFS)-based methods, and depth first search (DFS)-based methods. These two types of search algorithms tend to be inefficient, either generating too much load on the system [4], [5], or not meeting users' requirements [6]. Flooding, which belongs to BFS-based methods, is the default search algorithm for Gnutella network [7], [8]. By this method, the query source sends its query messages to all of its neighbors. When a node receives a query message, it first checks if it has the queried resource. If yes, it sends a

- T. Lin is with the Department of Electrical Engineering and Graduate Institute of Communication Engineering, National Taiwan University, BL626, No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan, ROC.

- .P. Lin and H. Wang are with the Graduate Institute of Communication Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan, ROC.

- C. Chen is with the Department of Electrical Engineering, National Taiwan University, No. 5, Lane 455, Sec. 6, Roosevelt Rd., Wunshan District, Taipei 116, Taiwan, ROC.

response back to the query source to indicate a query hit. Otherwise, it sends the query messages to all of its neighbors, except for the one the query message comes from. The drawback of flooding is the search cost. It produces considerable query messages even when the resource distribution is scarce. The search is especially inefficient when the target is far from the query source because the number of query messages would grow exponentially with the hop counts. Fig. 1 illustrates the operation of flooding. The link degree of each vertex in this graph is 4. If the network grows unlimited from the query source, the number of query messages generated by flooding at each hop would be 4, 12, 36, . . . , respectively. If the queried resource locates at one of the third neighbors, it takes 4þ12þ36¼52 query messages to get just one query hit.

On the other hand, random walk (RW) is a conservative search algorithm, which belongs to DFS-based methods [9], [10], [11], [12], [13]. By RW, the query source just sends one query message (walker) to one of its neighbors. If this neighbor does not own the queried resource, it keeps on sending the walker to one of its neighbors, except for the one the query message comes from, and thus, the search cost is reduced. The main drawback of RW is the long search time. Since RW only visits one node for each hop, the coverage of RW grows linearly with hop counts, which is slow compared with the exponential growth of the coverage of flooding. Moreover, the success rate of each query by RW is also low due to the same coverage issue. Increasing the number of walkers might help improve the search time and success rate, but the effect is limited due to the link degree and redundant path. As the example shown in Fig. 1, RW can only visit 12 vertices of second neighbors even when the number of walkers is set as 32. Certainly, the search is inefficient because 32 walkers only visit 12 vertices at the second hop.

LIN ET AL.: DYNAMIC SEARCH ALGORITHM IN UNSTRUCTURED PEER-TO-PEER NETWORKS
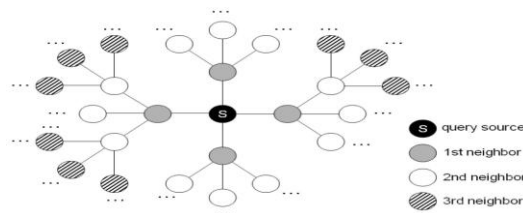
Fig. 1. A simple scenario of P2P network to demonstrate the operation of flooding and RW.

DS overcomes the disadvantages of flooding and RW and takes advantage of different contexts under which each search algorithm performs well. The operation of DS resembles flooding for the short-term search and RW for the long-term search. In order to analyze the performance of DS, we apply the random graphs as the models of network topologies and adopt the probability generating functions to model the link degree distribution [14]. We evaluate the performance of search algorithms in accordance with some performance metrics including the success rate, search time, number of query hits, number of query messages, query efficiency, and search efficiency [9], [15], [16]. Simulation experiments are performed in a dynamic P2P networking environment in order to collect convincing results for algorithm evaluations. The factors considered include the network topology, link degree distribution, peer's joining and leaving, and querying behavior as well as the activity of file sharing [10], [17], [18], [19]. Our dynamic network model is constructed based on these factors that strongly reflect the real measurement studies [17], [20], [21], [22]. Numerical results show that DS could provide a good tradeoff between search performance and cost. On average, DS performs about 25 times better than flooding and 58 times better than RW in power-law graphs, and about 186 times better than flooding and 120 times better than RW in bimodal topologies.The rest of this paper is organized as follows: Section 2 shows the related works about the search issue in unstructured P2P networks, followed by the detailed description of the proposed DS algorithm in Section 3. The performance analysis is given in Section 4. Numerical results and discussions are given in Section 5. Finally, the conclusion is presented in Section 6.

II.   RELATED WORKS

Flooding and RW are two typical examples of blind search algorithms by which query messages are sent to neighbors without any knowledge about the possible locations of the queried resources or any preference for the directions to send. Some other blind search algorithms include modified BFS (MBFS) [23], directed BFS [6], expanding ring [17], and random periodical flooding (RPF) [24]. These algorithms try to modify the operation of flooding to improve the efficiency. However, they still generate a large amount of query messages. Jiang et al. propose a LightFlood algorithm, which is a combination of the initial pure flooding and subsequent tree-based flooding [25], [26]. DS and LightFlood operate analogously, but DS avoids the extra cost to construct and maintain the treelike suboverlay.Knowledge-based search algorithms take advantage of the knowledge learned from previous search results and route query messages with different weights based on the knowl-edge. Thus, each node could relay query messages more intelligently. Some examples are adaptive probabilistic search (APS) [27], [28], biased RW [29], routing index (RI) [30], local indices [31], and intelligent search [32]. APS builds the knowledge with respect to each file based on the past experiences. RI classifies each document into some thematic categories and forwards query messages more intelligently based on the categories. The operation of local indices is similar to that of super-peer networks. Each node collects the file indices of peers within its predefined radius. If a search request is out of a node's knowledge, this node would perform a flooding search. The intelligent search uses a function to compute the similarity between a search query and recently answered requests. Nodes relay query messages based on the similarity. There are some other research works that focus on replicating a reference pointer to queried resources in order to improve the search time [33], [34].

III.   DYNAMIC SEARCH ALGORITHM

In this section, we provide the details of the proposed DS algorithm. Section 3.1 presents the operation of DS algorithm, and Section 3.2 provides the mechanism to combine DS with the knowledge-based search algorithms.

3.1   Operation of Dynamic Search Algorithm

DS is designed as a generalization of flooding, MBFS, and RW. There are two phases in DS. Each phase has a different searching strategy. The choice of search strategy at each phase depends on the relationship between the hop count h of query messages and the decision thresh-old n of DS.

3.1.1 Phase 1. When h _ n

At this phase, DS acts as flooding or MBFS. The number of neighbors that a query source sends the query messages to depends on the predefined transmission probability p. If the link degree of this query source is d, it would only send the query messages to d_p neighbors. When p is equal to 1, DS resembles flooding. Otherwise, it operates as MBFS with the transmission probability p.

3.1.2 Phase 2. When h > n

At this phase, the search strategy switches to RW. Each node that receives the query message would send the query message to one of its neighbors if it does not have the queried resource. Assume that the number of nodes visited by DS at hop $h\frac{1}{4}n$ is the coverage c_n, and then the operation of DS at that time can be regarded as RW with c_n walkers. However, there are some differences between DS and RW when we consider the whole operation. Consider the simple scenario shown in Fig. 1. Assume that the decision threshold n is set as 2. When h >2, DS performs the same as RW with $c_2\frac{1}{4}12$ walkers. Let us consider an RW search with $K\frac{1}{4}12$ walkers. At the first hop, the walkers only visit four nodes, but the cost is 12 messages.

```
Algorithm: The pseudo-code of dynamic search DS
Input: query source s, queried resource f, transmission probabil-
ity p
Output: the location information of f
DS(s, f, p)
/* the operation of s */
h ← 0
if (h <= n)
            h ← h + 1
            s choose p portion of its neighbors
            mᵢ carring h visits these chosen neighbors
elseif (h > n)
            h ← h + 1
            mᵢ carring h visits one neighbor of s
/* the operation of r */
foreach (r)
            if (r has the location information of f)
                    r returns the information to s
                    mᵢ stops
            elseif (h > TTL)
                    mᵢ stops
            elseif (h <= n)
                    h ←  h + 1
                    r choose p portion of its neighbors
                    mᵢ carring h visits these chosen neighbors
            elseif (h > n)
                    h ← h + 1
                    mᵢ carring h visits one neighbor of r
```

Fig. 2.The pseudocode of DS algorithm.
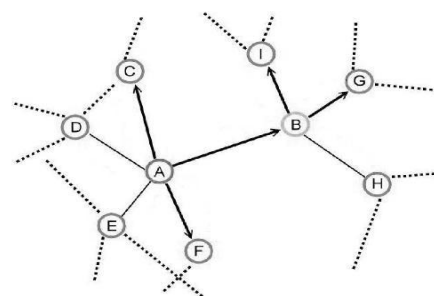
Fig. 3.Illustration for the operation of owledge-based DS algorithm.

RW would generate a large amount of redundant messages when K is set too large.Suppose that s is the query source, r is the vertex that receives the query message, f is the queried resource, m_i is the ith query message, and TTL is the time-to-live limitation. Fig. 2 shows the pseudocode of DS.In short, DS is designed to perform aggressively for the short-term search and conservatively for the long-term search. Obviously, the parameters n and p would affect the performance of DS. In Section 3.2, we will analyze the performance of DS andshow the effects of parameters n and p.

3.2 Knowledge-Based Dynamic Search

Some knowledge-based search algorithms, including APS, biased RW, RI, local indices, and intelligent search, are applicable to combine with our DS algorithm, and any training or caching operations are benefit from our DS algorithm as well. In this section, we present the generic scheme to incorporate these knowledge-based search algo-rithms with our DS algorithm. We construct the probabilistic function based on the information learned from the past experiences, with respect to each search target, search time, and local topology information. Thus, a node has more information to intelligently decide how many query mes-sages to send and to which peers these messages should be forwarded. Take APS as an example. The peer applying APS search builds a probability table for each neighbor and each object. It consistently refines its probability table by the search experiences. If a search query for some object delivers to a certain neighbor successfully, the probability entry corre-sponding to that neighbor and object is increased. If the search fails finally, it will decrease the probability entry. In accordance with APS, when a certain node receives a hit from peer i, it adds 10 points for the entry of peer i; if peer i fails to respond the hit to that node, the node subtracts 10 points for the entry of peer i.

Fig. 3 shows an example of knowledge-based DS algorithm. Node A initializes a search for a certain object. It makes its forwarding decision of which neighbors should be sent to in accordance with the probability table shown in Table 1. Assume the messages are sent to nodes B, C, and F. When node B receives the message, it checks its probability table shown in Table 1 and generates another two query messages to nodes I and G.

IV.   PERFORMANCE EVALUATION

In this section, we present the performance evaluation of DS. We apply Newman's random graph as the network topology, adopt the generation functions to model the link degree distribution [14], and analyze DS based on some performance metrics, including the success rate, search time, query hits, query messages, query efficiency, and search efficiency. The analysis by generating functions talks about a graph all of whose

parameters are exactly what they should be on an average random graph. Although the analysis using generating functions has appeared in many places by physicists, e.g., [10], it maybe not strict enough in the computer science context.Mihail et al. provide a strict analysis for RWs in power-law random graphs [35].

4.1 Network Model

First, we summarize Newman's work about the random graph. Let $G_0(x)$ be the generating function for the distribu-tion of the vertex degree k. $G_0(x)$ can be represented as

$$G_0(x) = \sum_{k=1}^{m} p_k x^k; \qquad (1)$$

wherep_k is the probability that a randomly chosen vertex in the graph has degree k, and m is the maximum degree.

TABLE 1

Node	C	D	E	B	F
Prob.	0.78	0.12	0.04	0.85	0.92

(a)

Node	G	H	I	-	-
Prob.	0.84	0.23	0.76	-	-

(b)

(a) Probability table for node A. (b) Probability table for node B.

Based on the generating function, the average degree of a randomly chosen vertex is given by

$$z_1 = \langle k \rangle = \sum_{k=1}^{m} k p_k = G_0^0(1): \qquad (2)$$

The average number of second neighbors is

$$z_2 = \frac{d}{dx} G_0(G_1(x)) \bigg|_{x=1} = G_0^0(1)G_1^0(1); \qquad (3)$$

where $G_1(x)$ is given by

$$G_1(x) = \frac{G_0^0(x)}{1(1)} = \frac{G_0^0(x)}{G_0^0} \qquad (4)$$

Due to the difficulties to correctly measure and sample the operational P2P networks, there are only limited real data about the topologies of such networks. In this paper, we will use the top two most common topologies, the power-law graphs and the bimodal topologies, to evaluate the search performance.

4.1.1 Power-Law Graphs

For the power-law random graph with the degree exponent _, p_k is proportional to k$^-$ [36]. That is,

$$p_k / k^- : \qquad (5)$$

According to [11], the following approximations for the power-law distribution are obtained:

$$G_0^0 ð1Þ \; ffi \; \frac{1}{2} \; ð1\_m^{2-}Þ \qquad\qquad ð6Þ$$

and

$$G_1^0 ð1Þ \; ffi \; G_0^0 ð1Þ \_ \frac{1}{3} \; \_{m}{3}\_ \qquad ; \qquad\qquad ð7Þ$$

assuming $2 <\_ < 3$.

4.1.2 Bimodal Topologies

For the bimodal network topology [12], [29], few ultra-peers are connected to a large number of nodes, and the rest have few neighbors. This assumption is regarded as realistic and followed by most papers such as [37] and [38]. The probability that a randomly chosen peer belongs to the

ultra-peers is denoted as p_{ultra}, and the probability that this peer belongs to the other part with few neighbors is thus
$p_{few} ¼ 1 \_p_{ultra}$. The degrees of the ultra-peers and the peers with few neighbors are denoted as k_{ultra} and k_{few}. Applying these parameters to (1), (2), (3), and (4), the

average number of neighbors at each hop for the bimodal topologies could be obtained.

4.2 Performance Analysis
4.2.1 Success Rate ðSRÞ

The success rate ðSRÞ is the probability that a query is successful, i.e., there is at least one query hit. Assume that the queried resources are uniformly distributed in the network with a replication ratio R. SR can be calculated as

$$SR ¼ 1 \_ð1 \_RÞ^C ; \qquad\qquad ð8Þ$$

where R is the replication ratio, and C is the coverage. This formula shows that SR highly depends on the coverage of the search algorithms. We use (8) to obtain an important performance metric, the search time ðSTÞ, in the following.

4.2.2 Search Time ðST Þ

To represent the capability of one search algorithm to find the queried resource in time with a given probability, we define the search time ðSTÞ as the time it takes to guarantee the query success with success rate requirement SR_{req}. ST represents the hop count that a search is successful with a probabilistic guarantee. Using (8), ST is obtained when the coverage C is equal to $\log_{ð1\_RÞ} ð1\_SR_{req}Þ$. For MBFS search algorithms, this situation occurs when

$$
\begin{aligned}
&p \quad G^0 \quad 1 \quad _p2 \; _G0 \quad 1 \; _G0 \quad 1 \quad _p3 \; _G0 \quad 1 \quad _G0 \quad 1 \quad 2 \\
&\quad\quad\quad\quad\quad\quad\quad\quad Þ \qquad\qquad ð \; Þ \\
&\_ \; 0 \qquad\qquad 0 \; \_ \; 1 ð Þ þ \; \_ \; 0 \quad \_ \; 1 ð Þ \\
&\qquad\qquad\qquad\qquad\qquad\qquad ST_{MBFS\_} \\
&\quad\quad {}^{ð Þ þ}_p ST_{MBFS}\_ð \; G^0 \quad 1 \quad\quad G^0 \quad 1 \quad 1\_ \qquad\qquad \_ \qquad 9 \\
&þ \_ \_ \qquad\qquad ð Þ \qquad\qquad ð \\
&\_þ \qquad\quad \_ \; 0 \_ \quad 1 \; Þ\_ \qquad\qquad\qquad ð \; Þ \\
&\quad ¼ \; {}^{\log}ð1\_RÞð^T\_{}^{SR}reqÞ^{:}
\end{aligned}
$$

Thus, ST for MBFS is

$$p\_ \; G^0 \quad 1 \qquad 1\_\log \qquad 1 \quad SR$$

$$S^{ST}MBF \quad {}^{log}p\_G_1{}^0ð \quad 1 ð Þ \_ \quad \frac{ð1\_R \quad req}{p \quad G} \quad Þ^ð \quad \_ \quad Þ \quad 1 :$$
$$¼ \quad \_ \quad \_ \quad {}_0{}^0ð1Þ \quad Þ \quad !$$
$$ð10Þ$$

ST of flooding is analog to that of MBFS with probability p ¼ 1.

The calculation of RW depends on the number of walkersk. When k is set as 1, ST for RW is obviously $\log_{ð1\_RÞ}ð1\_SR_{req}Þ$. When k is larger than 1, assume that

$$G_0{}^0ð1Þ\_G_1{}^0ð1Þ^{-t-1}\_ k \_ G_0{}^0ð1Þ\_G_1{}^0ð1Þ^{-t}; \quad ð11Þ$$

i.e., k is equal to or larger than the average number of the tthneighbors of the query source, and assume that theeffect of redundant paths can be neglected, and then the calculation for ST of RW can be expressed as

$$G_0^0 \quad 1 \quad {}_G0 \quad 1 \quad {}_G0 \quad 1 \quad {}_G0 \quad 1 \quad {}_G0 \quad 1 \quad t\_1$$
$$ð \quad Þ \quad ð \quad Þ þ \_ \_ \quad ð$$
$$0ð \quad Þ þ \quad 0 \quad \_ \quad 1 \quad \_þ \quad 0ð \quad Þ \_ \quad 1 \quad Þ \quad 12$$
$$SR \quad : \quad ð \quad Þ$$

$$þ \quad k\_ðST_{RW}\_tÞ ¼ \log_{ð1\_RÞ}ð1\_ \quad {}^-_{req}Þ \_$$

ST for RW is

$$t \quad 1 \quad i$$
$$\overline{\hspace{5cm}}$$

$$ST_{RW}¼ \quad {}^{log}ð1\_RÞð^1\_{}^{SR}reqÞ\_k \, {}^P_{i¼0} {}^-_0 \, {}^G_0{}^0ð1Þ\_{}_{G1}0_{ð1Þ}{}^- \quad þt: \quad ð13Þ$$

Now, we consider ST for DS. When the hop count h of the query message is smaller than or equal to the decision threshold n, ST of DS is equal to (10). When h is larger than n, the calculation can be expressed as

$$p \quad G^0 \quad 1 \quad p^2 \quad G^0 \quad 1 \quad G^0 \quad 1 \quad p^3 \quad G^0 \quad 1 \quad G^0 \quad 1 \quad 2$$
$$ð \quad Þ \quad ð \quad Þ \quad ð$$
$$\_ \quad ðð \quad Þ þ \quad \_ \quad (\quad \_ \quad 1ð \quad Þ þ \_ \quad 0 \quad \_ \quad 1 \quad Þ$$
$$p^n \quad G \quad 1 \quad G \quad {}^{n\_1}\_{pnG1} \quad \quad 14$$
$$0 \quad (\quad 1 \quad 0 \quad \_ \quad \_$$
$$\_0ð \quad Þ$$
$$þ \quad \_ \quad \_ \quad 1 \quad 0 \quad Þ \quad ð \quad Þ$$
$$\_{G}\_0 \quad n\_$$
$$þ \_ \quad {}^- \quad 1 \quad 1 \quad {}^{ST}DS \_ \quad {}_nð \quad Þ\_ \log^þ \quad {}^- \quad 1 \quad ð \quad {}^{SR}req \quad :$$
$$\_\_ \quad {}_1ð \quad Þ^- \quad ð \quad \_ \quad Þ ¼ \quad ð1\_RÞ^{ð-} \quad Þ$$

Therefore, ST for DS is

$$ST_{DS} = n + \frac{\log(1\_R)(1\_{SR}{}^{req})}{p^n\ G^0\ 1\quad G^0\ 1\ n\_1}$$

$$\frac{1}{\left[\begin{array}{c} G^{-}\ _1 \\ \_\ 0_p\ 0\ _n - 1 \\ \_\ \eth\ \TH\ \_ \\ \_\ p\ G\ 1\ 1\quad p\ G\ 1\ n\_1 \\ \_\ 1^0\ \eth{-}\TH\_\ \_\ \_\_\ _1{}^0\eth\ \TH \\ SR\ re \\ \_\ \log\ 1\ R\TH1\ \_\ \_\ q \\ \eth\_\ ^{\eth}\ \_\quad \TH \end{array}\right]} \qquad \eth15\TH$$

$$\text{ffi } n + \ _p n\_{G0}0_{\eth1\TH}\_{G1}0_{\eth1\TH}{}^{-n\_1}\ \_1:$$

We compare ST's for DS and RW with one walker. The improvement ratio is

$$\frac{ST_{RW}\_\ ST_{DS}}{ST_{RW}} \quad 1 \quad \frac{G_1^0\ \TH\ \frac{1}{\eth1}}{\text{ffi}\ \_\ G_0^0\ \TH\ \_p\_G_1^{0n}}\ \eth1\TH- \ : \quad 16 \quad \eth\ \TH$$

In (16), the last term on the right would significantly affect the performance improvement. ST of DS would be exponen-tially decreased with n, which can be expressed as O(1=n). Larger p would also affect the performance, but the effect is slow when compared with n. The extreme case of n is that it is set as TTL, i.e., DS performs as flooding or MBFS. In this case, ST would be the shortest, whereas it would also generate ahuge amount of query messages at the same time. The tradeoff between the search performance and the cost should be taken into consideration. In the following paragraphs, we further analyze the number of query hits and the number of query messages and further combine these metrics into the query efficiency and search efficiency.

4.2.3 Query Hits (QH)

The number of query hits highly depends on the coverage, i.e., the number of total visited nodes. Assume that the queried resources are uniformly distributed with the replication ratio R in the network, and the coverage is C. The number of query hits is R_C. The coverage C can be regarded as the summation of the coverage at each hop. Therefore, we first analyze the coverage C_h at the hth hop. Let V_h be the event that a vertex is visited at the hth hop. Suppose the probability that the vertex i is visited at the hth hop is $P_i(V_h)$. When the hop count h = 1, C_h is the expectation of thevertices that are visited at the first hop. When the hop count h is larger than 1, the calculation of C_h should preclude the event that the vertex has been visited in the previous hop. Therefore, the coverage C_h at the hth hop can be written as

$$C_h = \begin{cases} \sum_i^n 1\ P_i(V_h); & \text{for } h = 1; \\ \sum_{i=1}^n \left[1\ \prod_{j=1}\ P_i{}^{V_j}\right]\ P_i{}^{V_h}; & \text{for } h \geq 2; \end{cases} \qquad 17\ \eth\ \TH$$

whereN is the total number of vertices in the network. Next, we analyze the visiting probability $P_i(V_h)$ for flooding, MBFS, RW, and DS, respectively. First, we consider the flooding and MBFS cases. The visiting probability $P_i(V_h)$ of flooding or MBFS is

$$P_i(V_h) = p\ \_\ p_i\ G_0^0(1); \quad G_1^0\ 1\ ^{C_h}\_1\ ; \quad \text{for } h = 1;$$

$$1 \quad 1^- \quad p \quad p_i \qquad \qquad \text{for } h \qquad 2; \quad ð18Þ$$

$$\_ \quad \_ \quad \_ \quad \_ \qquad ð \, Þ^- \qquad \qquad \_$$

where p_i is the probability that vertex i is to be reached by certain edge. Aiello et al. [39] shows that p_i can be written as

$$p_i^{1\!\!4}P_{n\frac{1\!\!4m=i1\_}{}}^{m=i^{1-}} \qquad \qquad ð19Þ$$

$$\frac{}{i \quad 1}$$

where $\_$ is the power-law exponent, and m is the maximum degree.

When considering RW, we first calculate the probability that a vertex i is the candidate of RW, i.e.,

$$P_i ð R_h Þ \; ¼ \; \_1 \_ \quad 1 \quad \begin{matrix} G \\ p_i & {}^0_0 & 1 & ; \\ & {}^{ð\,Þ}p_i & G_1^0 & 1 & \_1 \end{matrix} \quad ; \quad \begin{matrix} {}^C h & \text{for } h & 1; \\ & \text{for } h \; ¼ \; 2: \end{matrix} \quad ð20Þ$$

$$\_ \quad \_ \quad \_ \quad \_ \qquad ð \_ \qquad \qquad \_$$

Then, the average number of candidates of RW at hop h is

$$r_h ¼ P_i ð R_h Þ : \qquad \qquad ð21Þ$$
$$\overset{n}{\underset{i¼1}{}}$$

Hence, the probability that vertex i is visited at hop h for RW is

$$\overset{X}{}$$
$$\begin{matrix} h & & i \end{matrix}$$
$$P_i ð V_h Þ \; ¼ P_i ð R_h Þ \_ \quad 1 \_ ð1 \_ \quad 1=r_h Þ^k \; ; \qquad ð22Þ$$

where k is the number of walkers.

The calculation of visiting probability $P_i ð V_h Þ$ for DS depends on the relation between h and n. When h $\_$ n, $P_i ð V_h Þ$ is given by (18). When h > n, (20), (21), and (22) are used to get $P_i ð V_h Þ$, where k in (22) is set as C_n, i.e., the coverage at the nth hop. Therefore, the visiting probability $P_i ð V_h Þ$ of DS is given by

$$P_i V_h \; \begin{matrix} & & & & & & \text{for } h \; ¼ \\ p & p_i & G_0^0 ð1Þ; & & {}^C h\_1 & 1; \\ 8 \; 1 \_ \quad \_ \_ \quad p & p_i & G_1^0 & 1 & ; & \text{for } 2 \quad h \quad n; \end{matrix}$$
$$ð \quad Þ ¼ > \_ \quad \_ \quad \_ \quad \_ \quad ð \, Þ \quad C_n \qquad \_ \quad \_$$
$$\begin{matrix} R^- & & 1=r \\ {}^< P_i & {}_h & 1 & 1 & {}^{\neg}_h & ; & \text{for } h > n: \\ > & ð & Þ\_{}^h \_ ð & \_ & Þ \; i \end{matrix}$$
$$:$$
$$\qquad \qquad \qquad ð23Þ$$

4.2.4 Query Messages ðQMÞ

When considering the flooding and MBFS cases, the query message e_h generated at hop h is given by

$$e_h \quad p \_ G_0^0 ð1Þ; \qquad \text{for } h \; ¼ \; 1; \qquad 24$$
$$\quad ¼ \_ p \_ G_1^0 ð1Þ \_ C_{h\_1}; \quad \text{for } h \_ 2: \qquad ð \quad Þ$$

When considering the RW case, the number of query messages for each hop keeps fixed as k, i.e., the number of walkers. Therefore, the total number of query messages for RW is k $\_$ TTL.

The calculation of query messages for DS depends on h and n. The query messages e_h generated at hop h for DS can

be written as

$$e_h = \begin{cases} pG^0_0 = 1; & \text{for } h = 1; \\ {}^{8}_{p\_G1}0 \ \eth_1 \Þ \ {}^{C}h\_1; & \text{for } 2 \_ h \_ n; \\ \eth \Þ \\ \frac{1}{4} {}^{<}C_n; \_ & \text{for } h > n: \end{cases} \qquad \eth25\Þ$$

:

4.2.5 Query Efficiency ðQEÞ

The number of query hits ðQHÞ and the number of query messages ðQMÞ are the well-known performance metrics for the evaluations of search algorithms. Generally speak-ing, the objective of search algorithms is to get the most query hits with the fewest query messages, but these two metrics often conflict with each other. Therefore, it requires a more objective metric to evaluate the search performance.

LIN ET AL.: DYNAMIC SEARCH ALGORITHM IN UNSTRUCTURED PEER-TO-PEER NETWORKS

659

We adopt the performance metrics proposed in [15], query efficiency ðQEÞ and search efficiency ðSEÞ, which consider both the search performance and the cost. The similar criterion can also be found in [9]. First, we calculate QE. In [15], QE is defined as

$$QE \frac{1}{4} \frac{\sum^{TTL}_{h=1} QH_h}{QM} \ \eth \Þ\_R \ ; \qquad \eth26\Þ$$

whereQHðhÞ is the query hits at the hth hop, QM is the total number of query messages generated during the query, and R is the replication ratio of the queried object. Since a search getting hits in a faster fashion delivers better users' experiences and should be gauged as the higher reputation, we modify (26) and show two types of QE's. QE$_1$is calculated as (26) shows, and QE$_2$penalizes searchresults coming from far away, i.e.,

$$QE_2 \ \frac{1}{4} \frac{\sum^{TTL}_{h=1} QH_h = h}{QM} \ {}^{\eth \Þ}\_R : \qquad \eth27\Þ$$

4.2.6 Search Efficiency ðSEÞ

The search efficiency ðSEÞ is proposed as a unified performance metric for search algorithms [15]. A similar criterion can be found in [9]. While the query efficiency QE does not consider the success rate SR, SE is defined as

$$SE \ \frac{1}{4} \frac{\sum^{TTL}_{h=1} QH_h = h \ SR}{QM} \ {}^{\eth \Þ}\_R \ ; \qquad \eth28\Þ$$

where QHðhÞ=h is the query hits in the hth hop weighted by the hop count, QM is the total number of query messages generated during the query, SR is the probability that the query is successful, i.e., there is at least one query hit, and R is the replication ratio of the queried object. Thus, the success rate SR is taken into consideration. Assume that the object is uniformly distributed in the network. Then, the query hit at the hth hop is equal to the multiplication of the coverage at the hth hop and the replication rate R. Therefore, (28) can be written as

$$\sum^{TTL}_{C_h} \_R = h \ 1 \_ \eth 1 \_ R \Þ^P \ h \ C_h$$

$$SE \; ¼ \; \frac{\displaystyle\sum_{P \; ¼}^{P} \; \frac{1 \quad C_h \; ¼ \; TT}{h \; 1 \quad L}{h}}{} \; _ \; \frac{1}{R} \; ; \quad ð29Þ$$

where C_h is the coverage at the hth hop, e_h is the query messages generated at the hth hop, and R is the replication ratio. We consider two types of SEs. SE_1 does not penalize search results coming from far away, i.e.,

$$SE_1 \; ¼ \; \frac{\displaystyle\sum_{P \; ¼}^{P} \; \frac{h \; 1 C_h \quad R \; 1}{e \; ¼ TTL}}{h}{} \; _ \; \frac{\frac{TT}{L}}{1 \quad R} \frac{TTL \, C_h}{h ¼ 1} \; _ \; ð \; \frac{1}{RÞ^P} \; ; \quad ð30Þ$$

and SE_2 is calculated as (29) shows.

4.3 Experimental Environment

We construct the experimental environment to evaluate the performance of the knowledge-based DS algorithm. For the network topology modeling, we model the P2P network as Gnutella to provide a network context in which peers can perform their intended activities. The measurements in [17]

and [20] have suggested that the topology of Gnutella network has the property of two-segment power-law link distribution. Thus, we construct a P2P network of 100,000 peers in our simulator, in which the link distribution follows the reported two-segment power law. We set the first power-law slope as 0.2316 and the second as 1.1373, which are similar to the ones used in [17]. The statistics result of the topology embedded in our simulator are that the maximum link degree is 632, mean is 11.73, and standard deviation is 17.09. Once the node (peer) degrees are chosen, we connect these peers randomly and reassure every peer is connected properly (each peer has at least one link).

For the object distribution of the network, we assume there are 100 distinct objects with replication ratio of R ¼ 1 percent; totally, there are 100,000 objects in thenetwork. The distribution of the 100,000 objects over the network follows the measurement characteristics reported in [21]. In addition, due to the dynamic environment— peers join and leave dynamically—described in the following section, the total number of objects available in the network will fluctuate according to the network size (number of online peers), but the replication ratio will roughly remain constant.

Our dynamic peer behavior modeling largely follows the proposed idea of the peer cycle [18], which includes joining, querying, idling, leaving, and joining again to form a cycle. The joining and leaving operations of peers (include idling) are inferred and then modeled by the uptime and session duration distributions measured in [21] and [22]. These measurement studies show similar results in the peer uptime distribution, where half of the peers have uptime percentage less than 10 percent and the best 20 percent of peers have 45 percent uptime or more. We use the log-quadratic distribution suggested in [22] to rebuild the uptime distribution, which is plotted in Fig. 4. However, for the session duration distribution, those two studies lead to different results. The median of session time in [22] is about 15 minutes, while it is 60 minutes in [21]. In our modeling, we choose the median session duration time to be

20 minutes.
By these two rebuilt distributions, we can generate a
probability model to decide when a peer should join or leave the network and how long it should continually be online. The basic rule to assign peers' attributes is that peers with higher link degrees are assigned to higher uptime percentages and longer session durations, and vice versa. With these conditions, we map a 2-hour-long dynamic join/ leave pattern for peers. On average, there are 10 peers joining or leaving simultaneously. Since the mean value of uptime distribution is about 18 percent, the resulting average number of online peers is 18,152. Moreover, the maximum number of online nodes is 24,218, while the minimum number is 4,886.
We model the dynamic querying model as Poisson distribution with the idle time _¼50 minutes; that is, each peer will initiate a search every 50 minutes on average. Since there is no direct measurement about the idle time, we just use an experiential value. The choice of this parameter is insensitive to our search performance
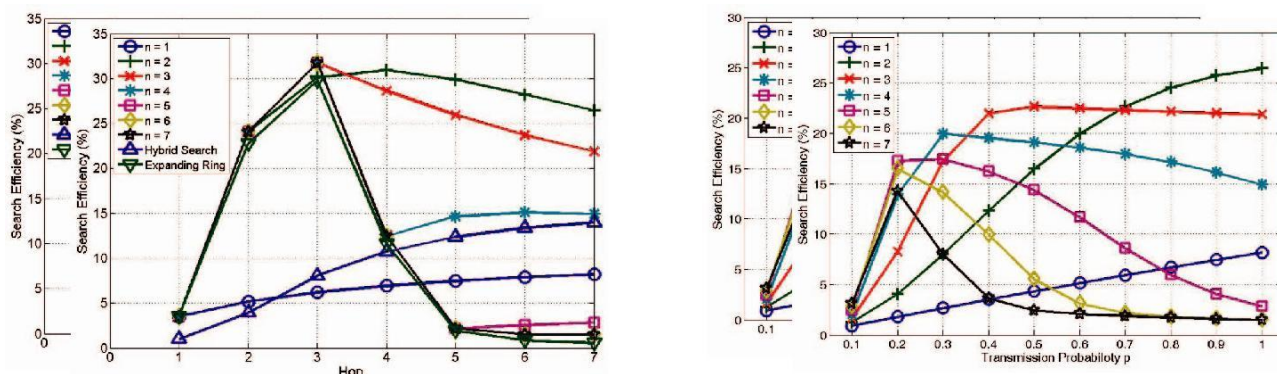
evaluation. With

Fig. 4. SE versus hop count when p is set as 1 and n is changed from 1 to 7. Power-law topology with N ¼ 10;000. When n is set as 2, DS gets the best performance for almost all hop counts.

Fig. 5. The effects of the parameters ðn; pÞ on the SE. Power-law topology with N ¼ 10;000. TTL ¼ 7. The best SE is obtained when ðn; pÞ is set as (2, 1).

the idle time of 50 minutes, there are thus about six queries or searches processing concurrently in the network on average. Totally, in this 2-hour simulation, we generate 43,632 search queries. Furthermore, for the query distribu-tion of search objects, we model it as zipf distribution with parameter a¼0:82, similar to the ones used in [17] and [27]. Finally, our simulator's central clock is triggered per second, which measures a hop for messaging passing and serves as a basic time unit for all peer operations.

V.   NUMERICAL RESULTS AND DISCUSSION

In this section, we show the numerical results of performance evaluation. We show the effectiveness of our DS algorithm and the effects of parameters n and p in Section 5.1. Then, the performance evaluation results of the knowledge-based DS algorithm are shown in Section 5.2.

5.1   Performance of Dynamic Search
5.1.1 Effects of Parameters n and p of DS

First, N is set as 10,000. Power-law topology is adopted and the exponent _ is set as 2.1, which is analog to the real-world situation [10]. Replication ratio R is set as 0.01 in this case. Fig. 4 illustrates how the decision threshold n of DS would affect the system performance. Due to space constraints, we only show the result when p is set as 1. The case n¼1 is analog to RW with K equal to the number of first neighbors, which is roughly 3.55 in this case. The case n¼7 is equal to the flooding. As this figure shows, DS with n¼7 sends the query messages aggressively in the first three hops and gets good SE. However, the perfor-mance degrades rapidly as the hop increases. This is because the cost grows exponentially with the path length between the query source and the target. On the contrary, SE of RW is better than that of the flooding when the hop is5 to 7. When n is set as 2, DS gets the best SE for almost all hop counts. This figure shows that a good choice of parameter n can help DS to take advantage of different contexts under which each search algorithm performs well.In order to obtain the best ðn; pÞ combination, we illustrate the ðn; p; SEÞ results in Fig. 5. Here, N is set as 10,000, R is set as 0.01, and TTL is set as 7. Under this

context, when p is large (0.7-1), setting n¼2 would get the best SE. Moreover, the best n value increases as p decreases, as Fig. 6 shows. For example, when p is set as 0.2, the best n would be 6 or 7. This is because when p is small, n should be increased to expand the coverage. On the contrary, n should be decreased to limit the growth of query messages when p is large. Therefore, the parameters n and p provide the tradeoff between the search performance and the cost. It shows the best SE is obtained when ðn; pÞ is set as (2, 1). Due to space constraints, the best parameters for other contexts are skipped in this paper, which can be found through similar operation.

5.1.2 Search Time

We show the numerical results of ST in Fig. 7. In this case, N is set as 10,000, R is set as 0.01, and TTL is set as 7.Similar results can be obtained when the parameters are set as other values. The walkers K for RW are set as 1 and 32. The decision thresholds n are set as 2, 3, and 7, and p is set as 1. TTL is set as 7 in this case, thus DS with n¼7 is equal to flooding. From Fig. 7, DS with large n always gets the short ST because it always covers more vertices. On the contrary, RW with K¼1 always gets the longest ST since its coverage is only incremental by one at each hop. When K is set as 32, its coverage is enlarged and ST can beimproved. However, DS still performs better than RW with 32 walkers even when n is set as only 2. Note that when n is set as 3, DS performs as well as that with n¼7, i.e., the
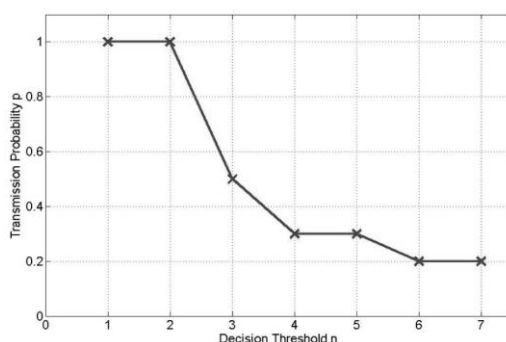
Fig. 6. The best ðn; pÞ combination when N is set as 10,000, R is set as

0.01, and TTL is set as 7.

LIN ET AL.: DYNAMIC SEARCH ALGORITHM IN UNSTRUCTURED PEER-TO-PEER NETWORKS 661
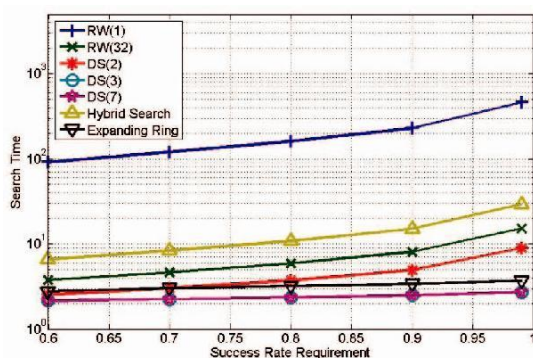
Fig. 7. ST versus SR requirement. R is set as 0.01 in this case. The walkers K for RW are set as 1 and 32, respectively. The n of DS are set as 2, 3, and 7, and p is set as 1. TTL is set as 7 in this case, thus the DS with n ¼ 7 is equal to flooding.
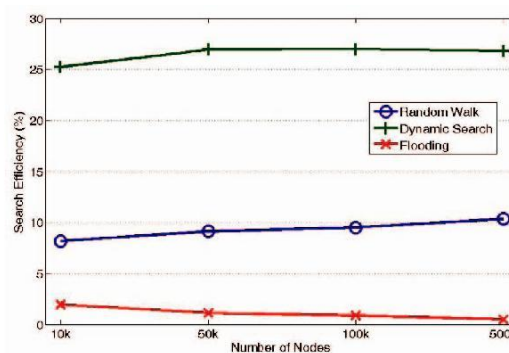
Fig. 8. Search efficiency for different number of nodes N in the network. This figure shows the scalability of the DS algorithm.

flooding, while does not generate as many query messages. In summary, DS with n¼2 and p¼1 would get the best SE and significantly improve ST in this case. Whileincreasing n to 3, although SE is a little degraded, the shortest ST is obtained.

5.1.3 Comparison with Other Advanced Search Algorithms

We also compare the performance of DS with that of other advanced search algorithms including Hybrid Search [12] and Expanding Ring [17]. The number of nodes N is set as 10,000. Power-law exponent _ is set as 2.1. Replication ratio R is set as 0.01 in this case. Fig. 4 shows SE's of these searchalgorithms. SE of Hybrid Search is analog to that of RW. They both increase slowly with hop counts. SE of Expanding Ring is analog to but a little worse than that of the flooding. This is because Expanding Ring would revisit the nodes it has already visited before. It would thus generate redundant messages. SE of DS is better than that of Hybrid Search and Expanding Ring for all hop counts.

Fig. 7 shows ST 's of these search algorithms. The operation of Hybrid Search is analog to that of RW with K ¼ $G^0_1ð1Þ$. Based on our simulation parameters, $G^0_1ð1Þ$ isroughly 16. Thus, ST of Hybrid Search is better than that of RW(1) but worse than that of RW(32). ST of Expanding Ring is almost one hop worse than that of the flooding. When the flooding reaches the second neighbors at the second hop, Expanding Ring just revisits the first neighbors and there is no increment in coverage. For SR requirement smaller than 0.7, ST of DS(2) is shorter than that of Expanding Ring, while ST of DS(2) would be longer than that of Expanding Ring for SR requirement larger than 0.7.

5.1.4 Scalability

In order to validate the scalability of our DS algorithm, we show the search efficiency for different number of nodes in Fig. 8. Nodes N are set as 10,000, 50,000, 100,000, and 500,000, respectively. The replication ratio R is set as 0.01, and TTL is set as 7. This figure shows that our DS algorithm always performs better than flooding and RW in spite of the number of nodes.

5.1.5 Performance under Various Network Topologies and Replication Ratios

Tables 2 and 3 show the search performance under power-law random graphs and bimodal topologies, respectively. The replication ratio R is set as 0.01 percent, 0.1 percent, and 1 percent, respectively. The performance metrics including the success rate ðSRÞ, search time ðSTÞ, number of query hits ðQHÞ, number of query messages ðQMÞ, query efficiency ðQEÞ, and search efficiency ðSEÞ are listed in these tables. Two types of QE's and SE's are shown. Ones without the penalty that the search results come from far away (QE_1 and SE_1), and others with the penalty (QE_2 and SE_2), as mentioned in Section 4.2. When considering $QE_1$and QE_2, RW performs the best because it covers the fewest redundant nodes. Although RW generates the fewest query messages, its SR, ST ,QH, and the resulting SE do not perform well. In most cases, DS can perform closely to the flooding search when considering SR and ST without generating as many query messages as flooding does. In summary, DS obtains satisfactory performances in spite of the number of nodes, the replication ratio, and the network topologies. On average, it performs about 25 times better than flooding and 58 times better than RW in power-law graphs, and about 186 times better than flooding and 120 times better than RW in bimodal topologies.

5.2 Performance of Knowledge-Based Dynamic Search

In this section, we evaluate the search performance in a network where every node is capable of building knowl-edge with respect to the target through some learning mechanisms. Any forwarding mechanism can improve the search performance by leveraging over the knowledge. For example, APS [27] uses the adaptive probability learning mechanism and adopts RW as the forwarding mechanism. Besides, other forwarding mechanisms, e.g., MBFS or our dynamic forwarding, are also applicable to this learning mechanism. In order to evaluate the search performance, we adopt APS learning mechanism to build the knowledge. APS learning builds a probability table for each neighbor and each object. When a query for certain object forwarding to a certain neighbor succeeds, the relative probability (or weight) of the entry for that neighbor and that object is

662IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 20, NO. 5, MAY 2009

TABLE 2a
Performance of Flooding in Power-Law Graphs

Size (N)	Replication Ratio (R) = 0.01%							
	Success rate (SR)	Search Time (ST)	Query Hits (QH)	Query Messages (QM)	Query Efficiency (QE_1)	Query Efficiency (QE_2)	Search Efficiency (SE_1)	Search Efficiency (SE_2)
10k	0.99	4.13	1.00	113k	0.088	1.25	0.087	1.24
50k	0.99	3.57	5.00	997k	0.050	1.13	0.050	1.12
100k	1.00	3.38	10.00	2561k	0.039	0.88	0.039	0.88
500k	1.00	3.03	50.00	23M	0.022	0.49	0.022	0.49
Size (N)	Replication Ratio (R) = 0.1%							
	Success rate (SR)	Search Time (ST)	Query Hits (QH)	Query Messages (QM)	Query Efficiency (QE_1)	Query Efficiency (QE_2)	Search Efficiency (SE_1)	Search Efficiency (SE_2)
10k	1.00	3.30	9.99	113k	0.088	1.96	0.088	1.96
50k	1.00	2.88	50.00	997k	0.050	1.13	0.050	1.13
100k	1.00	2.74	100.00	2560k	0.039	0.88	0.039	0.88
500k	1.00	2.48	500.00	23M	0.022	0.49	0.022	0.49
Size (N)	Replication Ratio (R) = 1%							
	Success rate (SR)	Search Time (ST)	Query Hits (QH)	Query Messages (QM)	Query Efficiency (QE_1)	Query Efficiency (QE_2)	Search Efficiency (SE_1)	Search Efficiency (SE_2)
10k	1.00	2.47	99.94	113k	0.088	1.96	0.088	1.96
50k	1.00	2.20	500.00	997k	0.050	1.13	0.050	1.13
100k	1.00	2.10	1K	2561k	0.039	0.88	0.039	0.88
500k	1.00	1.93	5K	23M	0.022	0.49	0.022	0.49

TABLE 2b
Performance of RW in Power-Law Graphs

Size (N)	Replication Ratio (R) = 0.01%							
	Success rate (SR)	Search Time (ST)	Query Hits (QH)	Query Messages (QM)	Query Efficiency (QE_1)	Query Efficiency (QE_2)	Search Efficiency (SE_1)	Search Efficiency (SE_2)
10k	0.0025	23k	0.0025	24.85	1.00	36.00	0.0025	0.09
50k	0.0028	23k	0.0028	28.18	1.00	35.71	0.0028	0.10
100k	0.0030	23k	0.0030	29.54	1.00	36.67	0.0030	0.11
500k	0.0032	23k	0.0033	32.53	1.00	37.50	0.0032	0.12
Size (N)	Replication Ratio (R) = 0.1%							
	Success rate (SR)	Search Time (ST)	Query Hits (QH)	Query Messages (QM)	Query Efficiency (QE_1)	Query Efficiency (QE_2)	Search Efficiency (SE_1)	Search Efficiency (SE_2)
10k	0.025	2k	0.025	24.85	1.00	36.40	0.025	0.91
50k	0.028	2k	0.028	28.18	1.00	36.79	0.028	1.03
100k	0.029	2k	0.030	29.54	1.00	37.24	0.029	1.08
500k	0.032	2k	0.033	32.53	1.00	37.19	0.032	1.19
Size (N)	Replication Ratio (R) = 1%							
	Success rate (SR)	Search Time (ST)	Query Hits (QH)	Query Messages (QM)	Query Efficiency (QE_1)	Query Efficiency (QE_2)	Search Efficiency (SE_1)	Search Efficiency (SE_2)
10k	0.22	229.11	0.25	24.85	1.00	37.23	0.22	8.19
50k	0.25	229.11	0.28	28.18	1.00	36.56	0.25	9.14
100k	0.26	229.11	0.30	29.54	1.00	36.62	0.26	9.52
500k	0.28	229.11	0.33	32.53	1.00	36.89	0.28	10.33

increased. Otherwise, it is decreased. Since the flooding forwards messages to all of the neighbors, the learning mechanism is useless for it, and so we do not evaluate flooding here. For the MBFS with APS learning, the transmission probability p is set as 0.2, which is chosen to keep the same amount of query messages as the other search algorithms. The initial walker for APS is 10, the same as [27].The experimental results for different search algorithms with the knowledge building mechanism are shown in Fig. 9. With APS knowledge building mechanism, all search algorithms perform much better than they do without

LIN ET AL.: DYNAMIC SEARCH ALGORITHM IN UNSTRUCTURED PEER-TO-PEER NETWORKS663

TABLE 2c
Performance of DS in Power-Law Graphs

Size (N)	Replication Ratio (R) = 0.01%							
	Success rate (SR)	Search Time (ST)	Query Hits (QH)	Query Messages (QM)	Query Efficiency (QE₁)	Query Efficiency (QE₂)	Search Efficiency (SE₁)	Search Efficiency (SE₂)
10k	0.75	4.46	1.00	15k	0.67	17.39	0.50	13.04
50k	0.99	3.57	5.00	91k	0.55	17.16	0.50	16.99
100k	1.00	3.38	10.00	200k	0.5	16.77	0.50	16.77
500k	1.00	3.03	50.00	1240k	0.40	16.16	0.40	16.16
Size (N)	Replication Ratio (R) = 0.1%							
	Success rate (SR)	Search Time (ST)	Query Hits (QH)	Query Messages (QM)	Query Efficiency (QE₁)	Query Efficiency (QE₂)	Search Efficiency (SE₁)	Search Efficiency (SE₂)
10k	0.89	4.40	2.21	2k	1.00	22.26	0.89	19.81
50k	1.00	2.88	7.79	8k	1.00	22.01	1.00	22.01
100k	1.00	2.74	13.49	14k	1.00	21.96	1.00	21.96
500k	1.00	2.48	48.93	49k	1.00	21.88	1.00	21.88
Size (N)	Replication Ratio (R) = 1%							
	Success rate (SR)	Search Time (ST)	Query Hits (QH)	Query Messages (QM)	Query Efficiency (QE₁)	Query Efficiency (QE₂)	Search Efficiency (SE₁)	Search Efficiency (SE₂)
10k	0.92	4.92	2.44	244.48	1.00	27.42	0.92	25.23
50k	0.99	2.96	4.85	485.15	1.00	27.21	0.99	26.94
100k	1.00	2.46	6.52	652.03	1.00	26.98	1.00	26.98
500k	1.00	1.93	12.96	1296.59	1.00	26.81	1.00	26.81

BLE 3a
Performance of Flooding in Bimodal Topologies

Size (N)	Replication Ratio (R) = 0.01%							
	Success rate (SR)	Search Time (ST)	Query Hits (QH)	Query Messages (QM)	Query Efficiency (QE₁)	Query Efficiency (QE₂)	Search Efficiency (SE₁)	Search Efficiency (SE₂)
10k	0.63	2.69	1.00	2497k	0.0040	0.14	0.0025	0.087
50k	0.99	2.69	5.00	12m	0.0042	0.13	0.0042	0.13
100k	1.00	2.69	10.00	24m	0.0042	0.12	0.0042	0.12
500k	1.00	2.69	50.00	125m	0.0040	0.11	0.0040	0.11
Size (N)	Replication Ratio (R) = 0.1%							
	Success rate (SR)	Search Time (ST)	Query Hits (QH)	Query Messages (QM)	Query Efficiency (QE₁)	Query Efficiency (QE₂)	Search Efficiency (SE₁)	Search Efficiency (SE₂)
10k	1.00	2.28	10.00	2497k	0.0040	0.14	0.0040	0.14
50k	1.00	2.28	50.00	12m	0.0042	0.13	0.0042	0.13
100k	1.00	2.28	100.00	24m	0.0042	0.12	0.0042	0.12
500k	1.00	2.28	500.00	125m	0.0040	0.11	0.0040	0.11
Size (N)	Replication Ratio (R) = 1%							
	Success rate (SR)	Search Time (ST)	Query Hits (QH)	Query Messages (QM)	Query Efficiency (QE₁)	Query Efficiency (QE₂)	Search Efficiency (SE₁)	Search Efficiency (SE₂)
10k	1.00	1.86	100.00	2497k	0.0040	0.14	0.0040	0.14
50k	1.00	1.86	500.00	12m	0.0042	0.13	0.0042	0.13
100k	1.00	1.86	1000.00	24m	0.0042	0.12	0.0042	0.12
500k	1.00	1.86	5000.00	125m	0.0040	0.11	0.0040	0.11

knowledge. Comparing these three search algorithms, for the case at h¼7, SE of DS is 24 percent better than that of RW, and 31 times better than that of MBFS. The outstanding performance results from the good tradeoff between the search performance and the cost.

VI. CONCLUSION

In this paper, we have proposed the DS algorithm, which is a generalization of the flooding, MBFS, and RW. DS overcomes the disadvantages of flooding and RW, and takes advantage of various contexts under which each search algorithm Authorized licensed use limited to: Shree MotilalKanhaiyalalFomra Institute. Downloaded on August 12,2010 at 08:33:11 UTC from IEEE Xplore. Restrictions apply.

664 IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 20, NO. 5, MAY 2009

TABLE 3b
Performance of RW in Bimodal Topologies

Size (N)	Replication Ratio (R) = 0.01%							
	Success rate (SR)	Search Time (ST)	Query Hits (QH)	Query Messages (QM)	Query Efficiency (QE$_1$)	Query Efficiency (QE$_2$)	Search Efficiency (SE$_1$)	Search Efficiency (SE$_2$)
10k	0.0014	23k	0.0014	13.99	1.00	37.14	0.0014	0.052
50k	0.0014	23k	0.0014	13.99	1.00	37.14	0.0014	0.052
100k	0.0014	23k	0.0014	13.99	1.00	37.14	0.0014	0.052
500k	0.0014	23k	0.0014	13.99	1.00	37.14	0.0014	0.052
Size (N)	Replication Ratio (R) = 0.1%							
	Success rate (SR)	Search Time (ST)	Query Hits (QH)	Query Messages (QM)	Query Efficiency (QE$_1$)	Query Efficiency (QE$_2$)	Search Efficiency (SE$_1$)	Search Efficiency (SE$_2$)
10k	0.014	2k	0.014	13.99	1.00	36.43	0.0014	0.51
50k	0.014	2k	0.014	13.99	1.00	36.43	0.0014	0.51
100k	0.014	2k	0.014	13.99	1.00	36.43	0.0014	0.51
500k	0.014	2k	0.014	13.99	1.00	36.43	0.0014	0.51
Size (N)	Replication Ratio (R) = 1%							
	Success rate (SR)	Search Time (ST)	Query Hits (QH)	Query Messages (QM)	Query Efficiency (QE$_1$)	Query Efficiency (QE$_2$)	Search Efficiency (SE$_1$)	Search Efficiency (SE$_2$)
10k	0.13	229.11	0.14	13.99	1.00	37.38	0.13	4.86
50k	0.13	229.11	0.14	13.99	1.00	37.38	0.13	4.86
100k	0.13	229.11	0.14	13.99	1.00	37.38	0.13	4.86
500k	0.13	229.11	0.14	13.99	1.00	37.38	0.13	4.86

TABLE 3c
Performance of DS in Bimodal Topologies

Size (N)	Replication Ratio (R) = 0.01%							
	Success rate (SR)	Search Time (ST)	Query Hits (QH)	Query Messages (QM)	Query Efficiency (QE$_1$)	Query Efficiency (QE$_2$)	Search Efficiency (SE$_1$)	Search Efficiency (SE$_2$)
10k	0.26	2.69	0.30	2964.41	1.00	26.15	0.26	6.80
50k	1.00	2.69	5.00	306k	0.16	16.28	0.16	16.28
100k	1.00	2.69	10.00	409k	0.24	19.02	0.24	19.02
500k	1.00	2.69	50.00	566k	0.88	21.33	0.88	21.33
Size (N)	Replication Ratio (R) = 0.1%							
	Success rate (SR)	Search Time (ST)	Query Hits (QH)	Query Messages (QM)	Query Efficiency (QE$_1$)	Query Efficiency (QE$_2$)	Search Efficiency (SE$_1$)	Search Efficiency (SE$_2$)
10k	0.95	2.28	2.96	2964.41	1.00	26.48	0.95	25.16
50k	0.95	2.28	2.99	2989.68	1.00	26.58	0.95	25.25
100k	0.95	2.28	2.99	2992.84	1.00	26.59	0.95	25.26
500k	0.95	2.28	3.00	2995.37	1.00	26.60	0.95	25.27
Size (N)	Replication Ratio (R) = 1%							
	Success rate (SR)	Search Time (ST)	Query Hits (QH)	Query Messages (QM)	Query Efficiency (QE$_1$)	Query Efficiency (QE$_2$)	Search Efficiency (SE$_1$)	Search Efficiency (SE$_2$)
10k	1.00	1.86	29.58	2964.41	1.00	26.54	1.00	26.54
50k	1.00	1.86	29.88	2989.68	1.00	26.59	1.00	26.59
100k	1.00	1.86	29.92	2992.84	1.00	26.59	1.00	26.59
500k	1.00	1.86	29.95	2995.37	1.00	26.60	1.00	26.60

performs well. It resembles flooding or MBFS for the short-term search and RW for the long-term search.
We analyze the performance of DS based on some metrics including the success rate, search time, number of query hits, number of query messages, query efficiency, and search efficiency. Numerical results show that proper setting of the

parameters of DS can obtain short search time and provide a good tradeoff between the search performance and cost. Under different contexts, DS always performs well. When combined with knowledge-based search algorithms, its search performances could be further improved. Authorized licensed use limited to: Shree MotilalKanhaiyalalFomra Institute. Downloaded on August 12,2010 at 08:33:11 UTC from IEEE Xplore. Restrictions apply.

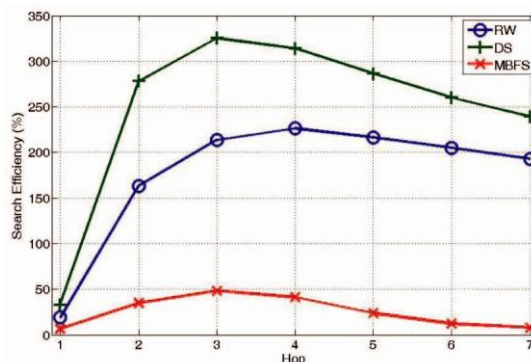LIN ET AL.: DYNAMIC SEARCH ALGORITHM IN UNSTRUCTURED PEER-TO-PEER NETWORKS
665

Fig. 9. Performance comparison when combined with the knowledge-based search mechanisms. DS always performs the best.

ACKNOWLEDGMENTS

REFERENCES

[1]   Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker, "Search and Replication in Unstructured Peer-to-Peer Networks," Proc. 16thAnn. Int'l Conf. Supercomputing (ICS '02), pp. 84-95, June 2002.

[2]   Z. Ge, D.R. Figueiredo, S. Jaiswal, J. Kurose, and D. Towsley, "Modeling Peer-Peer File Sharing Systems," Proc. IEEEINFOCOM '03, pp. 2188-2198, 2003.

[3]   K. Sripanidkulchai, The Popularity of Gnutella Queries and ItsImplications on Scalability, O'Reilly, www.openp2p.com, Feb. 2001.

[4]   M. Ripeanu, A. Iamnitchi, and I. Foster, "Mapping the Gnutella Network," IEEE Internet Computing, vol. 6, no. 1, pp. 50-56, Jan./Feb. 2002.

[5]   S. Saroiu, P.K. Gummadi, and S.D. Gribble, A Measurement Study ofPeer-to-Peer File Sharing Systems. MMCN, Jan. 2002.

[6]   J. Chu, K. Labonte, and B. Levine, "Availability and Locality Measurements of Peer-to-Peer File Systems," ITCom: Scalabilityand Traffic Control in IP Networks, July 2002.

[7]   V. Kalogeraki, D. Gunopulos, and D. Zeinalipour-Yazti, "A Local Search Mechanism for Peer-to-Peer Networks," Proc.ACM CIKM Int'l Conf. Information and Knowledge Management (CIKM '02), pp. 300-307, Nov. 2002.

[8]   Z. Zhuang, Y. Liu, L. Xiao, and L.M. Ni, "Hybrid Periodical Flooding in Unstructured Peer-to-Peer Networks," Proc. 32nd Int'lConf. Parallel Processing (ICPP '03), pp. 171-178, Oct. 2003.

[9]   S. Jiang, L. Guo, and X. Zhang, "LightFlood: An Efficient Flooding Scheme for File Search in Unstructured Peer-to-Peer Systems," Proc. 32nd Int'l Conf. Parallel Processing (ICPP '03), pp. 627-635, Oct. 2003.

[10]   S. Jiang, L. Guo, X. Zhang, and H. Wang, "LightFlood: Minimizing Redundant Messages and Maximizing Scope of Peer-to-Peer Search," IEEE Trans. Parallel and Distributed Systems, vol. 19, no. 5, pp. 601-614, May 2008.

[11]   D. Stutzbach, R. Rejaie, N. Duffield, S. Sen, and W. Willinger,   [27]   D. Tsoumakos and N. Roussopoulos, "Adaptive Probabilistic

[12]   "Sampling Techniques for Large, Dynamic Graphs," Proc. Ninth   Search for Peer-to-Peer Networks," Proc. Third Int'l Conf. Peer-to-

[13]   IEEE Global Internet Symp. (Global Internet '06), Apr. 2006.   Peer Computing (P2P '03), pp. 102-109, Sept. 2003.

[14]   A.H. Rasti, D. Stutzbach, and R. Rejaie, "On the Long-Term   [28]   D. Tsoumakos and N. Roussopoulos, "Analysis and Comparison

[15]   Evolution of the Two-Tier Gnutella Overlay," Proc. Ninth IEEEGlobal Internet Symp. (Global Internet '06), Apr. 2006.

[16]   D. Milojicic, V. Kalogeraki, R. Lukose, K. Nagaraja, J. Pruyne, B. Richard, S. Rollins, and Z. Xu, "Peer-to-Peer Computing," Technical Report HPL-2002-57, HP, 2002.

[17]   of P2P Search Methods," Technical Report CS-TR-4539, UMIACS-TR-2003-107, Dept. of Computer Science, Univ. of Maryland, 2003.

[18]   [29]Y. Chawathe, S. Ratnasamy, L. Breslau, N. Lanham, and S. Shenker, "Making Gnutella-Like P2P Systems Scalable," Proc.ACM SIGCOMM '03, pp. 407-418, Aug. 2003.

[19]   K. Sripanidkulchai, The Popularity of Gnutella Queries and Its[30] A. Crespo and H. Garcia-Molina, "Routing Indices for Peer-to-

[20]   Implications on Scalability, white paper, Carnegie Mellon Univ.,Feb. 2001.

[21]   M. Jovanovic, F. Annexstein, and K. Berman, "Scalability Issues in Large Peer-to-Peer Networks: A Case Study of Gnutella," technical report, Laboratory for Networks and Applied Graph Theory, Univ. of Cincinnati, 2001.

[22] B. Yang and H. Garcia-Molina, "Improving Search in Peer-to-Peer Networks," Proc. 22nd Int'l Conf. Distributed Computing Systems(ICDCS '02), pp. 5-14, July 2002.

[23] Peer Systems," Proc. 22nd Int'l Conf. Distributed Computing Systems(ICDCS '02), pp. 23-32, July 2002.

[24] B. Yang and H. Garcia-Molina, "Improving Search in Peer-to-Peer Networks," Proc. 22nd Int'l Conf. Distributed Computing Systems(ICDCS '02), pp. 5-14, July 2002.

[25] V. Kalogeraki, D. Gunopulos, and D. Zeinalipour-Yazti, "A Local Search Mechanism for Peer-to-Peer Networks," Proc. 11thInt'l Conf. Information and Knowledge Management (CIKM '02), pp. 300-307, Nov. 2002.

[26] G. Kan, "Gnutella," Peer-to-Peer Harnessing the Power of Disruptive[33] R.A. Ferreira, M.K. Ramanathan, A. Awan, A. Grama, and Technologies, O'Reilly, pp. 94-122, 2001.

[27] RFC-Gnutella 0.6, http://rfc-gnutella.sourceforge.net /developer/ testing/index.html, 2008.

[28] S. Jagannathan, "Search with Probabilistic Guarantees in Unstructured Peer-to-Peer Networks," Proc. Fifth IEEE Int'lConf. Peer-to-Peer Computing (P2P '05), pp. 165-172, Aug. 2005.

[29] C. Gkantsidis, M. Mihail, and A. Saberi, "Random Walks in [34] N. Sarshar, P.O. Boykin, and V.P. Roychowdhury, "Percolation

[30] Peer-to-Peer Networks," Proc. IEEE INFOCOM '04, pp. 120-130, 2004.

[31] L.A. Adamic, R.M. Lukose, A.R. Puniyani, and B.A. Huberman, "Search in Power-Law Networks," Physical Rev., E, vol. 64, 046135, 2001.

[32] L.A. Adamic, R.M. Lukose, and B.A. Huberman, "Local Search in Unstructured Networks," Handbook of Graphs and Networks.

[33] 295-317, Wiley-VCH, 2003.

[34] C. Gkantsidis, M. Mihail, and A. Saberi, "Hybrid Search Schemes for Unstructured Peer-to-Peer Networks," Proc. IEEEINFOCOM '05, pp. 1526-1537, 2005.

[35] N. Bisnik and A. Abouzeid, "Modeling and Analysis of Random Walk Search Algorithm in P2P Networks," Proc. SecondInt'l Workshop Hot Topics in Peer-to-Peer Systems (HOT-P2P '05),

[36] 95-103, 2005.

[37] M.E.J. Newman, S.H. Strogatz, and D.J. Watts, "Random Graphs with Arbitrary Degree Distribution and Their Applications," Physical Rev., E, vol. 64, 026118, 2001.

[38] H. Wang and T. Lin, "On Efficiency in Searching Networks," Proc.IEEE INFOCOM '05, pp. 1490-1501, 2005.

[39] P. Lin, T. Lin, and H. Wang, "Dynamic Search Algorithm in Unstructured Peer-to-Peer Networks," Proc. Global Telecomm. Conf.(GLOBECOM '06), Nov. 2006.

[40] Search in Power Law Networks: Making Unstructured Peer-to-Peer Networks Scalable," Proc. Fourth IEEE Int'l Conf. Peer-to-PeerComputing (P2P '04), pp. 2-9, Aug. 2004. M. Mihail, A. Saberi, and P. Tetali, "Random Walks with Lookahead in Power Law Random Graphs," Internet Mathematics, 2006.

[41] L.A. Adamic, "The Small World Web," Proc. Third European Conf.Digital Libraries (ECDL '99), pp. 443-452, 1999.

[42] S. Behnel and A. Buchmann, "Models and Languages for Overlay Networks," Proc. VLDB Workshop Databases, Information Systemsand Peer-to -Peer Computing (DBISP2P '05), Aug. 2005.

[43] S. Behnel and A. Buchmann, "Overlay Networks—Implementation by Specification," Proc. ACM/IFIP/USENIX Sixth Int'l MiddlewareConf. (Middleware), 2005.

[44] W. Aiello, F. Chung, and L. Lu, "A Random Graph Model for Massive Graphs," Proc. 32nd Ann. ACM Symp. Theory of Computing(STOC '00), pp. 171-180, 2000.

[45] Authorized licensed use limited to: Shree MotilalKanhaiyalalFomra Institute. Downloaded on August 12,2010 at 08:33:11 UTC from IEEE Xplore. Restrictions apply.

An Integrated Approach for Plagiarism Detection System

Shilpa[1], Mr. Manoj Challa[2]

[1,]CMRIT, BANGALORE,
[2] Assoc Professor, Dept of CSE,CMRIT, Bangalore,

ABSTRACT:

Nowadays as Internet is becoming the primary media for information access and nearly every information is available in the Internet. Therefore, it becomes easier to use another author's contents from the Internet without giving proper reference. Others ideas, solutions or expressions are representing as one's own original work is known as plagiarism. In this paper, we propose a framework which works by integrating several analytical procedures. Scholarly documents under investigation are segmented into logical tree-structured representation using a procedure called DSEGMENT. Statistical methods are utilized to assign numerical weights to structural components under a technique called C-WEIGHT. The top weighted components from the structural components are fed into plagiarism detection technique called Optimized Semantic Role Labeler (OSRL).This technique analyses and compares text based on the semantic allocation for each term inside the sentence. The absolute hash function method is used to map large data sets of variable length to smaller data of fixed length by generating hash key .In terms of Recall and Precision, this method outperforms the results with the existing plagiarism detection methods.

KEYWORDS: *Plagiarism detection, Scientific Publication, Semantic Argument, Semantic Role, Semantic Similarity, Natural language processing, hash map.*

I. INTRODUCTION

P lagiarism is defined as the appropriation or imitation of the methods, ideas and views of another writer and representation of them as one's original work[1]. Plagiarism in scientific publications has increased and one day you will see your published work is used in another publication yet without proper references and citation. Scientific publications in the same field normally share the same general information. Besides, each publication should express a particular message that contributes to that field. Diverse contributions can be made in diverse areas; for example, solving new problems, giving suggestions to existing problems, experimenting new methods, comparing current methods, giving enhancement to results. Such these contributions of others are considered their ideas and should be acknowledged when reported in a further research. The big challenge is to provide plagiarism detecting with proper method in order to improve the time required to check the results and percentage of finding results. Several plagiarism detection tools use character or string matching method to detect the plagiarized content. Most of the current software's and techniques are less effective in detecting a plagiarized text because these tools tend to compare the suspected text with original text using characters matching, some with chunks while others by words example MOSS[3],jplag[4]. This leads to rigorous search which takes a long time in the matching process. The matching algorithms are working depending on the text lexical structure rather than semantic structure. Therefore it becomes difficult to detect the text paraphrased semantically. One of the goals of this study is to propose new semantic techniques for plagiarism detection based on optimized Semantic Role Labeler. The proposed method does not analyze the content of a text document as text syntax only, but also captures the underlying semantic meaning in terms of the relationships among its terms.

In this paper we propose a new plagiarism detection technique based on analytical procedures and Optimized Semantic Role Labeler. In analytical procedures we highlight the similar sections by giving the statistical weights. Optimized Semantic Role Labeler can help in detecting sophisticated obfuscation such as copy paste, anonyms, renaming, paraphrasing, splitting, merging, and changing the sentence from active to passive voice and vice versa in the highlighted document. The absolute hash function is used to map large data sets of variable length to smaller data of fixed length by generating hash key.

By using the absolute hash function we can reduce the retrieving time in the list. The remaining section of the paper is structured as follows; Section II discusses the related literature survey on plagiarism detection. Section III discusses the framework for plagiarism detection using OSRL. Section IV explores the experimental analysis. In section V, we discuss the results and the potential market of OSRL, and we draw a conclusion and future of this work in the last section.

II.   RELATED WORK

The above section gives a overview of the existing plagiarism methods. Many tools works on finger print based method, String matching method. Fingerprint method [5] [9] works by creating fingerprint for each document in the collection. It gives the detailed information about the number of terms per line, number of keywords, and number of unique terms. Mostly rely on the use of K- grams (Manuel et al. 2006) because the process of fingerprinting divides the document into grams of certain length k.Different plagiarism tools have been surveyed in multiple works. Many online solutions are established today; for example, CrossCheck,Turnitin,SafeAssign,EVE2, WCopyFind ,Viber, Scriptum, PlagiarismDetect, SCAM ,CHECK , PPChecker , SNITCH , Ferret and others. Several studies have reported that using plagiarism checker tools in academia is effective in reducing the problem, and discouraging the students to commit plagiarism . Besides, anti-plagiarism tools help to educate students and authors in different disciplines about plagiarism .

JPlag [4], Sherlock are the examples of String Matching- based algorithm. Many String Matching algorithm works based on tokenization, where each file is replaced by predefined tokens. One of the problem of JPlag[4] is that files must be parse to be included in the comparison for plagiarism because of this some similar files that are not parsed to be missed. Unlike JPlag, in Sherlock files do not have to parse to be included in the comparison and also displays quick visualization. The time complexity and space complexity is more in the above method and also not checking at sentence level. DANIEL R WHITE and MIKE S JOY[6] proposed a method by comparing suspected documents at sentence level there by detects paraphrasing, reordering, merging ,splitting but fails to analyze semantically and also speed is less compared to other software like CopyCatch.
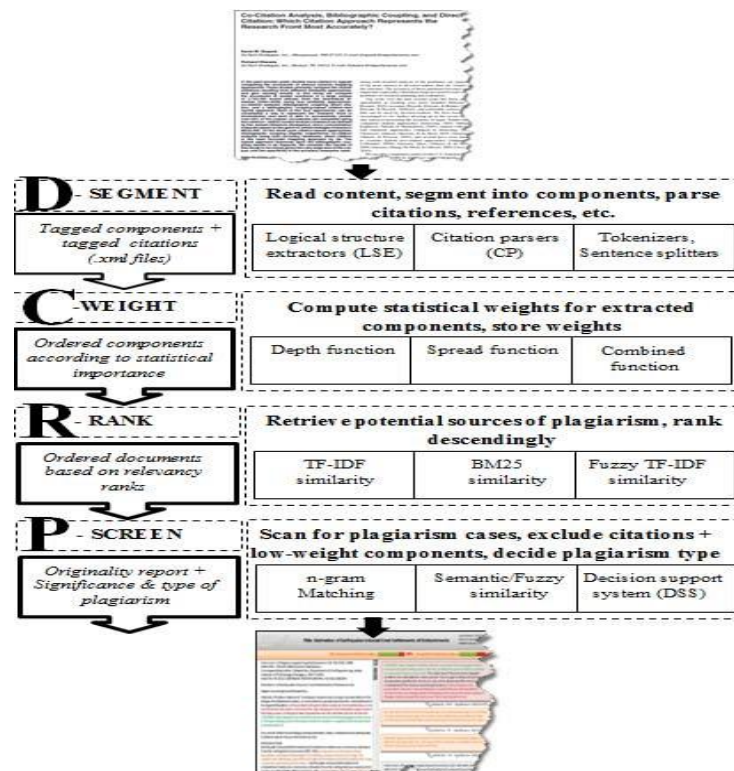
Figure 1: Existing plagiarism detection framework Iplag

Iplag[7] gives a accurate judgment about type of plagiarism. It works by combining several analytical procedures. Documents under investigation are segmented into D segment [7].Then weights are given to D-segment called as C-weight. Relevance Ranking(R-Ranking) and plagiarism Screening approach [7] (P-screen) is used. The architecture for Iplag which is useful for existing plagiarism detection systems is shown in

figure1.This Iplag acts as a framework and can be applied to any of the existing plagiarism detection technique.Plagiarism Detection Scheme Based on Semantic Role Labeling [2] analyses and compares text based on the semantic allocation for each term inside the sentence. It identifies and label arguments in a text. The idea behind SRL is that the sentence level semantic analysis of text determines the object and subject of a text. For calculating the semantic labeling, to retrieve semantic arguments and labels in large document will take large amount of time and space. The architecture as shown below

Figure 2: Structure of Semantic Role Labeler

III. FRAMEWORK FOR PLAGIARISM DETECTION USING OSRL

This paper proposes a framework that is very accurate and reliable in reporting plagiarism in scientific publications. Checking plagiarism manually requires an Expert. An Expert must have idea about the contributions in different parts of paper like discussions, conclusi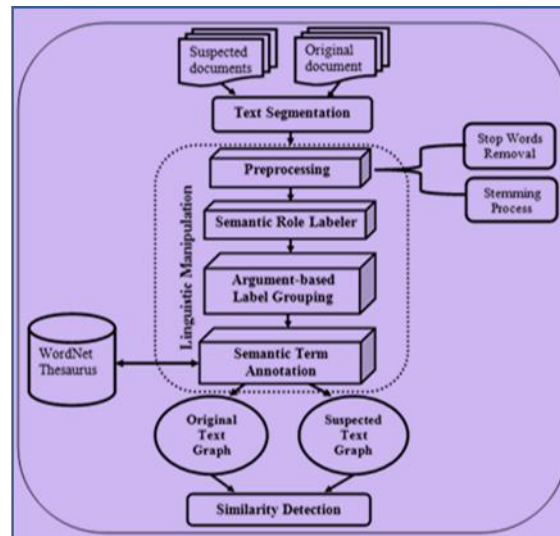ons and any suspected methods, techniques, ideas that might be taken from somewhere else. The experts may not pay attention to every statement but only sentences that convey original ideas and ignores important parts. The OSRL is used for sentence level semantic analysis of the text that determines the object and subject of a text. It relays on the characterization of events such as determining "when", "where ", how: who" did,"what" to "whom". The predicate of a clause (normally a verb) establishes "what "took place and other parts of the sentence express the other arguments of the sentence. The main goal of OSRL is to identify what semantic relation holds among a predicate and its associate participants with relations drawn from a predefined list of possible semantic roles for that predicate.

3.1.Document Segmentation [D-segment]

The framework begins by splitting the scientific publications into several meaningful components [7] i.e.
• document->sections->paragraphs
• document->paragraphs->sentences
• document->topics->paragraphs

In this framework, the logical structural organizations of scientific publications were used in which it includes most of the following categories: Title, Owner, Abstract, Introduction, Literature survey (related previous works, etc.), Evaluation, Acknowledgements, and References. These categories are called generic classes. Further, generic classes involve different structural components such as head titles, paragraphs, tables, tablecaptions, equations, figure captions; etc, Structural components are further segmented into sentences.

3.2.Component Weighting: C-WEIGHT

By assigning numerical weights to structural component .we can show how important that component is to the article. Depth and Spread functions are statistical measures and are useful for detecting results. Spread of a term t defines the number of structural components that contain the term. Depth of a term t refers to the frequency of the term in a class (unlike normal term frequency which considers the number of the term's occurrences in the whole document).The weight w of a structural component c in a generic class Gc can be obtained by clubbing Spread and Depth ,that can be expressed as follows:

$$w(c) = \sum_{t \in c} spread(t) \times \frac{tf_{t,}G_c}{\max t f_{G_c}}$$ (1)

Where spread (t) is the number of components that has t, $tf_{t,}Gc$ is the frequency of t in the generic class Gc which has t, and $\max tf_{Gc}$ is the maximum frequency occurs in that class. Select the top weighted components that are having more weight.
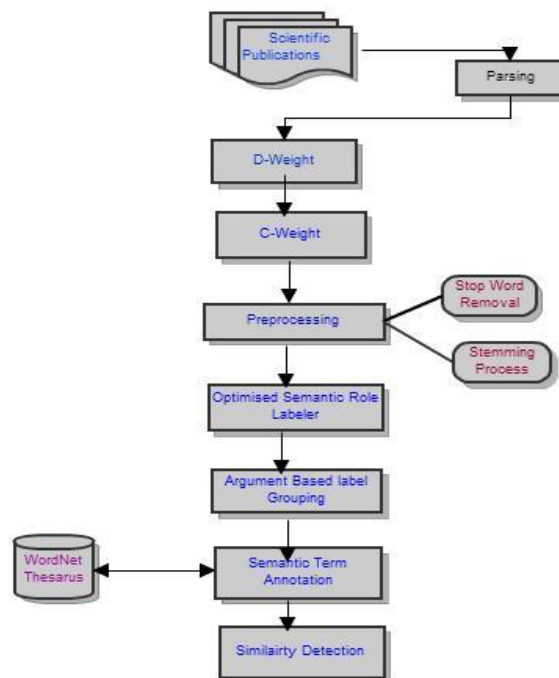
Figure 3 .Presents a Plagiarism Detection Framework procedure

3.3.OSRL for plagiarism detection

In this method, we take top weighted components of the suspected documents for pre-processing steps .i.e. stop words removal and stemming. Then, OSRL was used to transform the sentences into arguments based on location for each term in the sentences. The verbs play an important role in every sentence for processing, and the analysis of the sentences. All the arguments extracted from the text were grouped in the nodes according to the argument type. Each group was named by the argument name. This step is called Argument Label Group (ALG). Then, we extracted all the concepts for each term in the argument groups using WordNet thesaurus. This step is called Semantic Term Annotation (STA).

3.4.Similarity Detection

Create a hash values for the labels created from the OSRL. Similarity Detection between original and suspected document is done by first extracting the hash values of the suspected document and then check that with hash values of original documents. If two is matching then start finding similarity of labels with respect to hash values which are matched.

Plagiarist can be shown through the following example:James gave Rohan the parcel (original sentence).
The parcel was given to Rohan by James (suspected sentence)

By using OSRL the produced arguments are:
James gave Rohan the parcel
Output:

OSRL   Charnaik

James	giver [A0]	(S1 | S |NP (NNP James)|
gave	V:give	|VP (VBD gave)
Rohan	entity given to [A2]	(NP (NNP Rohan | |
the	thing given[A1]	(NP (DT the |
parcel		(NN parcel | | |
.		(. .)))

Figure 4. Analysis for original sentence using OSRL

The parcel was given to Rohan by James
Output:

		OSRL	*Charnaik*
The		(S1 | S (NP (DT The |	
Parcel	thing given [A1]	(NN parcel))	
Was		(VP (AUX was)	
Given	V:give	(VP (VBN given)	
To		(PP (T0 to)	
Rohan	entity given to [A2]	)))	(NP (NNP Rohan
By	giver[A0]	(PP (IN by)	
James		))))	(NP (NNP James)
.		(. .)))	

Figure 5. Analysis for suspected sentence using OSRL

Fig 4 and 5 explains how the suspected sentence analyzed using OSRL. The structure of two sentences may different if synonyms and antonyms or active and passive are used. But they might be same through semantically. The OSRL captures the arguments for a sentence in spite of changing the places for the labels inside the sentences. This method of analyzing the sentence supports our proposed method in detecting plagiarism if comparison is applied based on the arguments of the sentence using OSRL.

3.5. Experimental Design and Dataset

The experiments were performed on 100 suspected documents, each plagiarized from one or more original documents. In this point, Original and tokenized suspected documents were analyzed by sentence-based similarity. Sentences in suspected documents were compared with each sentence in the candidate documents according to the arguments of the sentences. We not only detect the arrangement similarity between sentences, but also possible semantic similarity between two sentences. Similarity detection was conducted by comparing the original Topic Signature terms and suspected Topic Signature terms. If the two terms were found to be identical, we went directly to the argument label groups that contained these terms, and then determined the label group where they belonged, thus determining the possible sentences that may be plagiarized. This step compared the arguments of possible sentences that had been plagiarized with the corresponding arguments in original sentences. The argument label group gave way to the main arguments and each argument inside the group quickly take to the possible plagiarized sentence. Some parameter play a crucial role in the similarity calculation, such as the number of matched arguments and number of arguments which exist in the sentences. The first variable determines the similar arguments between the suspected document and original document

while the second variable determines the argument that does not exist in the sentences. The similarity between the arguments of the suspected document and original document was calculated according to Jaccard coefficient measure [8] which is well-known as famous similarity measure between two sets. Jaccard coefficient [8] defined as a following equation:

$$Similarity\ C_i(ArgS_j, ArgS_k) = \frac{C(ArgS_j) \cap C(ArgS_k)}{C(ArgS_j) \cup C(ArgS_k)} \quad (2)$$

Where,

$C(ArgS_j)$=concepts of the argument sentence in the suspected document; $C_i(ArgS_k)$ =concepts of the argument sentence in the original document;

We then calculated the similarity between the suspected document and original document based on the following equation:

TABLE 1.
EVALUATION MEASURE OF THE PROPOSED METHOD

Number of documents	Recall	Precision	F-measure
1000	0.828237	0.655606	0.761109

$$\text{Total Similarity (Doc1, Doc2)} = \sum_{i-Ll}\sum_{\substack{j-lm \\ k-Ln}} SimC_i(ArgS_j, ArgS_k) \quad (3)$$

Where,

$SimC_i(ArgS_j, ArgS_k)$ is similarity between Arguments sentence j in suspected document containing concept i Arguments sentence k in original document containing concept   i , l = no. of concepts,   m= no. of Arguments sentence in suspected document, n = no. of Arguments sentence in original document.

IV.   RESULTS AND DISCUSSION

Our technique was tested according to the group of 100 documents .Those suspected documents were plagiarized with different ways of plagiarism such as simple copy and paste, changing some terms with their corresponding synonyms, and modifying the structure of the sentences (paraphrasing). We provided three general testing parameters that are commonly used in plagiarism detection as following

$$Recall = \frac{Number\ of\ Detected\ Arguments}{Total\ Number\ of\ Arguments} \quad (4)$$

$$Precision = \frac{Number\ of\ Plagiarized\ Arguments}{Number\ of\ Detected\ Arguments} \quad (5)$$

$$F\text{-}Measure = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (6)$$

The proposed method is evaluated and compared with Latent Semantic Analysis technique. The results are shown below
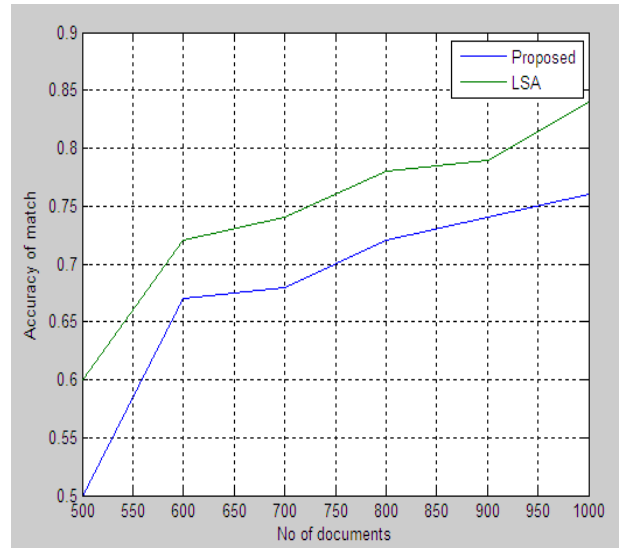
Figure 6. Comparison of two techniques in terms of accuracy.
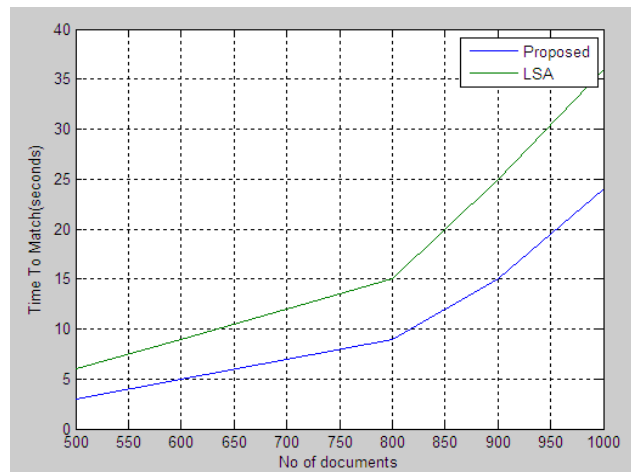
Figure 7. Comparison of two techniques in terms of time required for plagiarism detection.
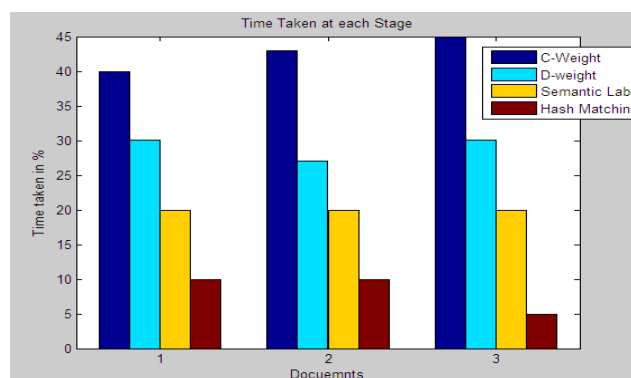
Figure 8. Time taken in each modules present in OSRL technique.

Fig. 6 and 7 demonstrates the comparison between OSRL and Latent Semantic Analysis in terms of accuracy and time required for Plagiarism Detection respectively. Fig 8 demonstrates the time required in each module which are present in OSRL technique. We found that all the scores that were obtained by our proposed method have good results than other method. The results from the comparison show that the proposed method achieved better results in terms of accuracy and time required for Detection.

V.   CONCLUSIONS AND FUTURE WORK

Optimized Semantic role labeling can be used for plagiarism detection by extracting argument of sentences and comparing the arguments. Tests were carried out against 100 dataset for plagiarism detection. The proposed methods were found to achieve better performance compared to Latent Semantic Analysis [9]. Our future work is to improve its efficiency and time complexity by introducing Hadoop Map Reduce Technique. As of now our detection technique works for mono lingual documents. We can extend it for multi lingual documents.

REFERENCES

[1]   www.academia.edu/689297/Source_Code_Plagiarism_-a_Student_Perspective
[2]   Naomie Salim, Ahmed Hamza Osman, "Plagiarism Detection Scheme Based on Semantic Role Labeling", International conference march 2012.
[3]   Http://theory.stanford.edu/aiken/moss/, 2005.
[4]   Lutz Prechelt , Guido Malpohl ,"JPlag: Finding plagiarisms among a set of programs" March , 2000 , http://wwwipd.ira.uka.de/EIR/
[5]   http://www.dcs.warwick.ac.uk/report/pdfs/cs-rr-440.pdf
[6]   DANIEL R. WHITE and MIKE S. JOY," Sentence-Based Natural Language Plagiarism Detection",IEEE ACM-TRANSACTION August 23, 2005
[7]   Salha Alzahrani , Naomie Salim , Ajith Abraham," iPlag: Intelligent Plagiarism Reasoner in Scientific Publications",International Conference,2011.
[8]   P. Jaccard, "Etude comparative de la distribution florale dans une portion des Alpes et des Jura," Bulletin de la Society Vaudoise des Sciences Naturelles, vol. 37, pp. 547–579, 1901.
[9]   Cosma, Georgina and Joy, Mike. (2012) An approach to source-code plagiarism detection and investigation using latent semantic analysis. IEEE Transactions on Computers, Vol.61 (No.3). pp. 379-394. ISSN 0018-9340

Mrs. Shilpa is presently doing Master of Technology in Computer science and Engineering at CMR Institute of Technology, Bangalore, Karnataka. She received her Bachelor of Engineering degree in Information Science and Engineering from Kalpatharu Institute of Technology, Tiptur, Karnataka in the year 2007.

Mr. Manoj Challa is pursuing Ph.D(CSE) in S.V.University, Tirupati, India. He completed his M.E(CSE) from Hindustan College of Engineering, Tamil Nadu in 2003. He is presently working as Associate Professor, CMR Institute of Technology, Bangalore. He presented nearly 18 papers in national and international conferences. His research areas include Artificial intelligence and computer networks.

Emergence of Hop Integrity in Computer Networks with Algorithms and Description of Protocols

Ashis Saklani

Assistant Professor Dept. Of Computer Science H.N.B Garhwal Central University Srinagar Garhwal Uttarakhand

ABSTRACT

A computer network is said to provide hop integrity iff when any router p in the network receives a message m supposedly from an adjacent router q, then p can check that m was indeed sent by q, was not modified after it was sent, and was not a replay of an old message sent from q to p. In this paper, we describe three protocols that can be added to the routers in a computer network so that the network can provide hop integrity. These three protocols are a secret exchange protocol, a weak integrity protocol, and a strong integrity protocol. All three protocols are stateless, require small overhead, and do not constrain the network protocol in the routers in any way.

KEYWORDS *: Security, Hop Integrity, Secret Exchange Protocol, Hop Integrity Protocol, Weak Integrity Protocol.*

I.   INTRODUCTION

Most computer networks suffer from the following security problem: in a typical network, an adversary, that has an access to the network, can insert new messages, modify current messages, or replay old messages in the network. In many cases, the inserted, modified, or replayed messages can go undetected for some time until they cause severe damage to the network. More importantly, the physical location in the network where the adversary inserts new messages, modifies current messages, or replays old messages may never be determined. Two well-known examples of such attacks in networks that support the Internet Protocol (or IP, for short) and the Transmission Control Protocol (or TCP, for short) are as follows.

II.   SMURF ATTACK:

In an IP network, any computer can send a "ping" message to any other computer which replies by sending back a "pong" message to the first computer as required by Internet Control Message Protocol (or ICMP, for short) [14]. The ultimate destination in the pong message is the same as the original source in the ping message. An adversary can utilize these messages to attack a computer d in such a network as follows. First, the adversary inserts into the network a ping message whose original source is computer d and whose ultimate destination is a multicast address for every computer in the network. Second, a copy of the inserted ping message is sent to every computer in the network. Third, every computer in the network replies to its ping message by sending a pong message to computer d. Thus, computer d is flooded by pong messages that it did not requested.

III.   SYN ATTACK:

To establish a TCP connection between two computers c and d, one of the two computers c sends a "SYN" message to the other computer d. When d receives the SYN message, it reserves some of its resources for the expected connection and sends a "SYN-ACK" message to c. When c receives the SYN-ACK message, it replies by sending back an "ACK" message to d. If d receives the ACK message, the connection is fully established and the two computers can start exchanging their data messages over the established connection. On the other hand, if d does not receive the ACK message for a specified time period of T seconds after it has sent the SYN-ACK message, d discards the partially established connection and releases all the resources reserved for that connection. The net effect of this scenario is that computer d has lost some of its resources for T seconds. An adversary can take advantage of such a scenario to attack computer d as follows [1, 18].

First, the adversary inserts into the network successive waves of SYN messages whose original sources are different (so that these messages cannot be easily detected and filtered out from the

network) and whose ultimate destination is d. Second, d receives the SYN messages, reserves its resources for the expected connections, replies by sending SYN-ACK messages, then waits for the corresponding ACK messages which will never arrive. Third, the net effect of each wave of inserted SYN messages is that computer d loses all its resources for T seconds. In these (and other [7]) types of attacks, an adversary inserts into the network messages with wrong original sources. These messages are accepted by unsuspecting routers and routed toward the computer under attack. To counter these attacks, each router p in the network should route a received m only after it checks that the original source in m is a computer adjacent to p or m is forwarded to p by an adjacent router q. Performing the first check is straightforward, whereas performing the second check requires special protocols between adjacent routers. In this paper, we present a suite of protocols that provide hop integrity between adjacent routers: whenever a router p receives a message m from an adjacent router q, p can detect whether m was indeed sent by q or it was modified or replayed by an adversary that operates between p and q. It is instructive to compare hop integrity with secure routing [2, 11, 17], ingress filtering [4], and IPsec [8]. In secure routing, for example [2], [11], and [17], the routing update messages that routers exchange are authenticated. This authentication ensures that every routing update message, that is modified or replayed, is detected and discarded. By contrast, hop integrity ensures that all messages (whether data or routing update messages), that are modified or replayed, are detected and discarded. Using ingress filtering [4], each router on the network boundary checks whether the recorded source in each received message is consistent with where the router received the message from. If the message source is consistent, the router forwards the message as usual. Otherwise, the router discards the message. Thus, ingress filtering detects messages whose recorded sources are modified (to hide the true sources of these messages), provided that these modifications occur at the network boundary. Messages whose recorded sources are modified between adjacent routers in the middle of the network will not be detected by ingress filtering, but will be detected and discarded by hop integrity. The hop integrity protocol suite in this paper and the IPsec protocol suite presented in [8], [9], [10], [12], and [13] are both intended to provide security at the IP layer. Nevertheless, these two protocol suites provide different, and somewhat complementary, services. On one hand, the hop integrity protocols are to be executed at all routers in a network, and they provide a minimum level of security for all communications between adjacent routers in that network. On the other hand, the IPsec protocols are to be executed at selected pairs of computers in the network,and they provide sophisticated levels of security for the communications between these selected computer pairs. Clearly, one can envision networks where the hop integrity protocol suite and the IPsec protocol suite are both supported. Next, we describe the concept of hop integrity in some detail. 2. Hop Integrity Protocols

A network consists of computers connected to subnetworks. (Examples of subnetworks are local area networks, telephone lines, and satellite links.) Two computers in a network are called adjacent iff both computers are connected to the same subnetwork. Two adjacent computers in a network can exchange messages over any common subnetwork to which they are both connected. The computers in a network are classified into hosts and routers. For simplicity, we assume that each host in a network is connected to one subnetwork, and each router is connected to two or more subnetworks. A message m is transmitted from a computer s to a faraway computer d in the same network as follows. First, message m is transmitted in one hop from computer s to a router r.1 adjacent to s. Second, message m is transmitted in one hop from router r.1 to router r.2 adjacent to r.1, and so on. Finally, message m is transmitted in one hop from a router r.n that is adjacent to computer d to computer d. A network is said to provide hop integrity iff the following two conditions hold for every pair of adjacent routers p and q in the network. i. Detection of Message Modification: Whenever router p receives a message m over the subnetwork connecting routers p and q, p can determine correctly whether message m was modified by an adversary after it was sent by q and before it was received by p. ii. Detection of Message Replay: Whenever router p receives a message m over the subnetwork connecting routers p and q, and determines that message m was not modified, then p can determine correctly whether message m is another copy of a message that is received earlier by p. For a network to provide hop integrity, two "thin" protocol layers need to be added to the protocol stack in each router in the network. As discussed in [3] and [16], the protocol stack of each router (or host) in a network consists of four protocol layers; they are (from bottom to top) the subnetwork layer, the network layer, the transport layer, and the application layer. The two thin layers that need to be added to this protocol stack are the secret
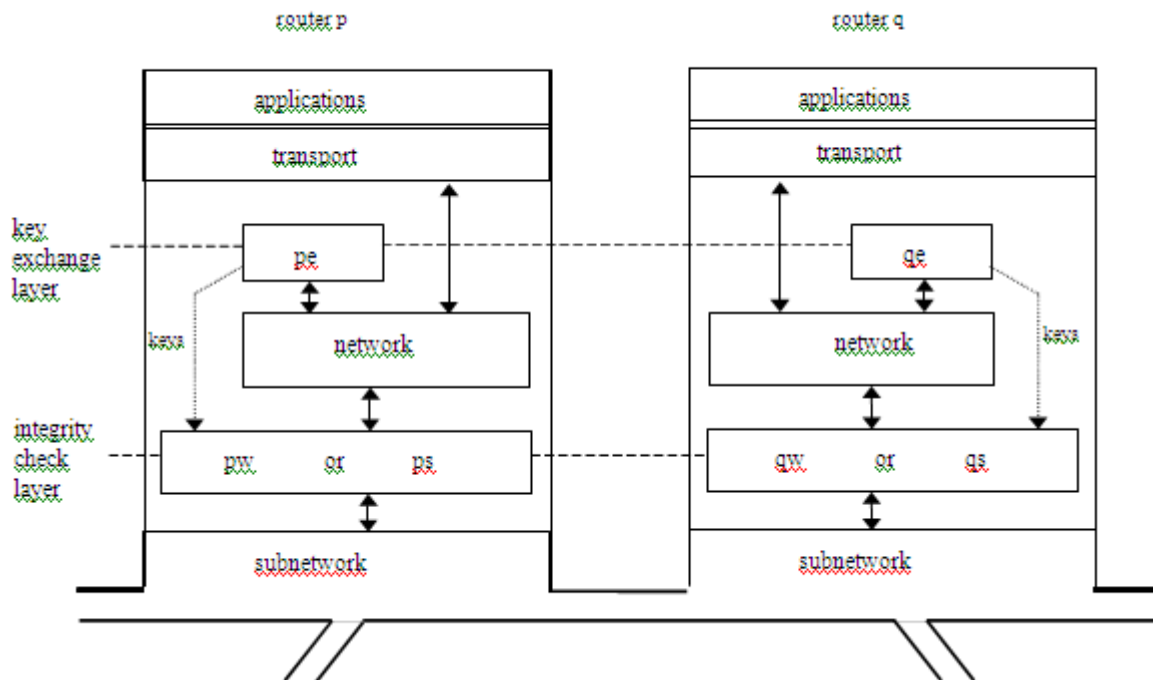
Figure 1.   Protocol stack for achieving hop integrity.

xchange layer and the integrity check layer. The secret exchange layer is added above the network layer (and below the transport layer), and the integrity check layer is placed below the network layer (and above the subnetwork layer). The function of the secret exchange layer is to allow adjacent routers to periodically generate and exchange (and so share) new secrets. The exchanged secrets are made available to the integrity check layer which uses them to compute and verify the integrity check for every data message transmitted between the adjacent routers. Figure 1 shows the protocol stacks in two adjacent routers p and q. The secret exchange layer consists of the two processes pe and qe in routers p and q, respectively. The integrity check layer has two versions: weak and strong. The weak version consists of the two processes pw and qw in routers p and q, respectively. This version can detect message modification, but not message replay. The strong version of the integrity check layer consists of the two processes ps and qs in routers p and q, respectively. This version can detect both message modification and message replay. Next, we explain how hop integrity, along with ingress filtering, can be used to prevent smurf and SYN attacks (which are described in the Introduction). Recall that in smurf and SYN attacks, an adversary inserts into the network ping and SYN messages with wrong original sources. These forged messages can be inserted either through a boundary router or between two routers in the middle of the network. Ingress filtering (which is usually

installed in boundary routers [4]) will detect the forged messages if they are inserted through a boundary router because the recorded sources in these messages would be inconsistent with the hosts from which these messages are received. However, ingress filtering may fail in detecting forged messages if these messages are inserted between two routers in the middle of the network. For example, an adversary can log into any host located between two routers p and q, and use this host to insert forged messages toward router p, pretending that these messages are sent by router q. The real source of these messages can not be determined by router p because router p cannot decide whether these messages are sent by router q or by some host between p and q. However, if hop integrity is installed between the two routers p and q, then the (weak or strong) integrity check layer in router p concludes that the forged messages have been modified after being sent by router q (although they are actually inserted by the adversary and not sent by router q), and so it discards them. Smurf and SYN attacks can also be launched by replaying old messages. For example, the adversary can log into any host located between two routers p and q. When the adversary spots some passing legitimate ping or SYN message being sent from q to p, it keeps a copy of the passing message. At a later time, the adversary can replay these copied messages over and over to launch a smurf or SYN attack. Hop integrity can defeat this attack as follows. If hop integrity is installed between the two routers p and q, then the strong integrity check layer in

router p can detect the replayed messages and discard them. In the next three sections, we describe in some detail the protocols in the secret exchange layer and in the two versions of the integrity check layer. The first protocol between processes pe and qe is discussed in Section 3. The second protocol between processes pw and qw is discussed in Section 4. The third protocol between processes ps and qs is discussed in Section 5. These three protocols are described using a variation of the Abstract Protocol Notation presented in [5]. In this notation, each process in a protocol is defined by a set of inputs, a set of variables, and a set of actions. For example, in a protocol consisting of processes px and qx, process px can be defined as follows. process px

```
inp   <name of input>        :      <type of input>

      ...

      <name of input>        :      <type of input>
var   <name of variable>   :      <type of variable>

      ...

      <name of variable>   :      <type of variable>
begin
      <action>
[]    <action>

      ...

[]    <action>
end
```

Comments can be added anywhere in a process definition; each comment is placed between the two brackets { and }. The inputs of process px can be read but not updated by the actions of process px. Thus, the value of each input of px is either fixed or is updated by another process outside the protocol consisting of px and qx. The variables of process px can be read and updated by the actions of process px. Each <action> of process px is of the form: <guard>□ □ □ □ □ □ <statement> The <guard> of an action of px is either a <boolean expression> or a <receive> statement of the form: rcv <message> from qx The <statement> of an action of px is a sequence of skip, <assignment>, <send>, or <selection> statements. An <assignment> statement is of the form: <variable of px> := <expression> A <send> statement is of the form: send <message> to qx A <selection> statement is of the form:

```
         <boolean expression>        →      <statement>
         ...
if       <boolean expression>
                                      →      <statement>
[]
fi
```

Executing an action consists of executing the statement of this action. Executing the actions (of differen processes) in a protocol proceeds according to the following three rules. First, an action is executed only when its guard is true. Second, the actions in a protocol are executed one at a time. Third, an action whose guard is continuously true is eventually executed. Executing an action of process px can cause a message to be sent to process qx. There are two channels between the two processes: one is from px to qx, and the other is from qx to px. Each sent message from px to qx remains in the channel from px to qx until it is eventually received by process qx or is lost. Messages that reside simultaneously in a channel form a sequence <m.1; m.2; □ ; m.n> in accordance with the order in which they have been sent. The head message in the sequence, m.1, is the earliest sent, and the tail message in the sequence, m.n, is the latest sent. The messages are to be received in the same order in which they were sent. We assume that an adversary exists between processes px and qx, and that this adversary can perform the following three types of actions to disrupt the communications between px and qx. First, the adversary can perform a message loss action where it discards the head message from one of the two channels between px and qx. Second, the adversary can perform a message modification action where it arbitrarily modifies the contents of the head message in one of the two channels between px and qx.

Third, the adversary can perform a message replay action where it replaces the head message in one of the two channels by a message that was sent previously. For simplicity, we assume that each head message in one of the two channels between px and qx is affected by at most one adversary action. 3. The Secret Exchange Protocol In the secret exchange protocol, the two processes pe and qe maintain two shared secrets sp and sq. Secret sp is used by router p to compute the integrity check for each data message sent by p to router q, and it is also used by router q to verify the integrity check for each data message received by q from router p. Similarly, secret sq is used by q to compute the integrity checks for data messages sent to p, and it is used by p to verify the integrity checks for data messages received from q. As part of maintaining the two secrets sp and sq, processes pe and qe need to change these secrets periodically, say every te hours, for some chosen value te. Process pe is to initiate the change of secret sq, and process qe is to initiate the change of secret sp. Processes pe and qe each has a public key and a private key that they use to encrypt and decrypt the messages that carry the new secrets between pe and qe. A public key is known to all processes (in the same layer), whereas a private key is known only to its owner process. The public and private

$$
\textbf{var} \quad
\begin{array}{ll}
sp : & \textbf{integer} \\
sq : & \textbf{array}\ [0 .. 1]\ \textbf{of} \\
 & \textbf{integer}
\end{array}
$$

keys of process pe are named B p and R p respectively; similarly the public and private keys of process qe are named B q and R q respectively. For process pe to change secret sq, the following four steps need to be performed. First, pe generates a new sq, and encrypts the concatenation of the old sq and the new sq using qe's public key B q, and sends the result in a rqst message to qe. Second, when qe receives the rqst message, it decrypts the message contents using its private key R q and obtains the old sq and the new sq. Then, qe checks that its current sq equals the old sq from the rqst message, and installs the new sq as its current sq, and sends a rply message containing the encryption of the new sq using pe's public key B p. Third, pe waits until it receives a rply message from qe containing the new sq encrypted using B p. Receiving this rply message indicates that qe has received the rqst message and has accepted the new sq. Fourth, if pe sends the rqst message to qe but does not receive the rply message from qe for some tr seconds, indicating that either the rqst message or the rply message was lost before it was received, then pe resends the rqst message to qe. Thus tr is an upper bound on the round trip time between pe and qe. Note that the old secret (along with the new secret) is included in each rqst message and the new secret is included in each rply message to ensure that if an adversary modifies or replays rqst or rply messages, then each of these messages is detected and discarded by its receiving process (whether pe or qe). Process pe has two variables sp and sq declared as follows.

Similarly, process qe has an integer variable sq and an array variable sp. In process pe, variable sp is used for storing the secret sp, variable sq[0] is used for storing the old sq, and variable sq[1] is used for storing the new sq. The assertion sq[0] $\neq$ sq[1] indicates that process pe has generated and sent the new secret sq, and that qe may not have received it yet. The assertion sq[0] = sq[1] indicates that qe has already received and accepted the new secret sq. Initially, sq[0] in pe = sq[1] in pe = sq in qe, and sp[0] in qe = sp[1] in qe = sp in pe. Process pe can be defined as follows. (Process qe can be defined in the same way except that each occurrence of R p in pe is replaced by an occurrence of R q in qe, each occurrence of B q in pe is replaced by an occurrence of Bp in qe, each occurrence of sp in pe is replaced by an occurrence of sq in qe, and each occurrence of sq[0] or sq[1] in pe is replaced by an occurrence of sp[0] or sp[1], respectively, in qe.) process pe

```
    if sq[1] = d →   sq[0] := sq[1]
    [] sq[1] ≠ d →   {detect adversary} skip
    fi
timeout
    sq[0] ≠ sq[1] ∧
    (tr seconds passed since rqst message sent last) →
        e := NCR(B _, (sq[0]; sq[1]));
        send rqst(e) to qe
```

end The four actions of process pe use three functions NEWSCR, NCR, and DCR defined as follows. Function NEWSCR takes no arguments, and when invoked, it returns a fresh secret that is different from any secret that was returned in the past. Function NCR is an encryption function that takes two arguments, a key and a data item, and returns the encryption of the data item using the key. For example, execution of the statement e := NCR(B q, (sq[0]; sq[1])) causes the concatenation of sq[0] and sq[1] to be encrypted using the public key B q, and the result to be stored in variable e. Function DCR is a decryption function that takes two arguments, a key

and an encrypted data item, and returns the decryption of the data item using the key. For example, execution of the statement d := DCR(R p, e)

```
inp   R_p     : integer        {private key of pe}
```

causes the (encrypted) data item e to be decrypted using the private key R p, and the result to be stored in variable d. As another example, consider the statement (d, e) := DCR(R p, e) This statement indicates that the value of e is the encryption of the concatenation of two values (v 0; v 1) using key R p. Thus, executing this statement causes e to be decrypted using key R p, and the resulting first value v to be stored in variable d, and the resulting second value v1 to be stored in variable e. A proof of the correctness of the secret exchange protocol is presented in the full version of the paper [6]. 4. The Weak Integrity Protocol The main idea of the weak integrity protocol is simple. Consider the case where a data(t) message, with t being the message text, is generated at a source src then transmitted through a sequence of adjacent routers r.1, r.2, □ , r.n to a destination dst. When data(t) reaches the first router r.1, r.1 computes a digest d for the message as follows: d := MD(t; scr) where MD is the message digest function, (t; scr) is the concatenation of the message text t and the shared secret scr between r.1 and r.2 (provided by the secret exchange protocol in r.1). Then, r.1 adds d to the message before transmitting the resulting data(t, d) message to router r.2. When the second router r.2 receives the data(t, d) message, r.2 computes the message digest using the secret shared between r.1 and r.2 (provided by the secret exchange process in r.2), and checks whether the result equals d. If they are unequal, then r.2 concludes that the received message has been modified, discards it, and reports an adversary. If they are equal, then r.2 concludes that the received message has not been modified and proceeds to prepare the message for transmission to the next router r.3. Preparing the message for transmission to r.3 consists of computing d using the shared secret between r.2 and r.3 and storing the result in field d of the data(t, d) message. When the last router r.n receives the data(t, d) message, it computes the message digest using the shared secret between r.(n-1) and r.n and checks whether the result equals d. If they are unequal, r.n discards the message and reports an adversary. Otherwise, r.n sends the data(t) message to its destination dst. Note that this protocol detects and discards every modified message. More importantly, it also determines the location where each message modification has occurred. Process pw in the weak integrity protocol has two inputs sp and sq that pw reads but never updates. These two inputs in process pw are also variables in process pe,and pe updates them periodically, as discussed in the previous section. Process pw can be defined as follows. (Process qw is defined in the same way except that each occurrence of p, q, pw, qw, sp, and sq is replaced by an occurrence of q, p, qw, pw, sq, and sp, respectively.) process pw

```
inp   sp          : integer
      sq          : array [0 .. 1] of integer
var   t, d        : integer
begin
    rcv data(t, d) from qw →
        if MD(t; sq[0]) = d ∨ MD(t; sq[1]) = d →
            {defined later} RTMSG
        [] MD(t; sq[0]) ≠ d ∧ MD(t; sq[1]) ≠ d →
            {report adversary} skip
        fi
[]  true →
        {p receives data(t, d) from router other than q}
        {and checks that its message digest is correct}
        RTMSG
[]  true →
        {either p receives data(t) from an adjacent}
        {host or p generates the text t for the next}
        {data message}
        RTMSG
end
```

end In the first action of process pw, if pw receives a data(t, d) message from qw while sq[0]□ □ □ sq[1], then pw cannot determine beforehand whether qw computed d using sq[0] or using sq[1]. In this case, pw needs to compute two message digests using both sq[0] and sq[1] respectively, and compare the two digests with d. If

either digest equals d, then pw accepts the message. Otherwise, pw discards the message and reports the detection of an adversary. The three actions of process pw use two functions named MD and NXT, and one statement named RTMSG. Function MD takes one argument, namely the concatenation of the text of a message and the appropriate secret, and computes a digest for that argument. Function NXT takes one argument, namely the text of a message (which we assume includes the message header), and computes the next router to which the message should be forwarded. Statement RTMSG is defined as follows.

```
if NXT(t) = p →            {accept message} skip
[] NXT(t) = q →            d := MD(t; sp);
                           send data(t, d) to qw
[] NXT(t) ≠ p ∧ NXT(t) ≠ q →
     {compute d as the message digest of}
     {the concatenation of t and the secret}
     {for sending data to NXT(t); forward}
     {data(t, d) to router NXT(t)} skip
fi
```

A proof of the correctness of the weak integrity protocol is presented in the full version of the paper [6]. 5. The Strong Integrity Protocol The weak integrity protocol in the previous section can detect message modification but not message replay. In this section, we discuss how to strengthen this protocol to make it detect message replay as well. We present the strong integrity protocol in two steps. First, we present a protocol that uses "soft sequence numbers" to detect and discard replayed data messages. Second, we show how to combine this protocol with the weak integrity protocol (in the previous section) to form the strong integrity protocol. Consider a protocol that consists of two processes u and v. Process u continuously sends data messages to process v. Assume that there is an adversary that attempts to disrupt the communication between u and v by inserting (i.e. replaying) old messages in the message stream from u to v. In order to overcome this adversary, process u attaches an integer sequence number s to every data message sent to process v. To keep track of the sequence numbers, process u maintains a variable nxt that stores the sequence number of the next data message to be sent by u and process v maintains a variable exp that stores the sequence number of the next data message to be received by v. To send the next data(s) message, process u assigns s the current value of variable nxt, then increments nxt by one. When process v receives a data(s) message, v compares its variable exp with s. If $exp \leq s$, then q accepts the received data(s) message and assigns exp the value s + 1; otherwise v discards the data(s) message. Correctness of this protocol is based on the observation that the predicate $exp \leq nxt$ holds at each (reachable) state of the protocol. However, if due to some fault (for example an accidental resetting of the values of variable nxt) the value of exp becomes much larger than value of nxt, then all the data messages that u sends from this point on will be wrongly discarded by v until nxt becomes equal to exp. Next, we describe how to modify this protocol such that the number of data(s) messages, that can be wrongly discarded when the synchronization between u and v is lost due to some fault, is at most N, for some chosen integer N that is much larger than one. The modification consists of adding to process v two variables c and cmax, whose values are in the range 0..N- 1. When process v receives a data(s) message, v compares the values of c and cmax. If $c \leq cmax$, then process v increments c by one (mod N) and proceeds as before (namely either accepts the data(s) message if $exp \leq s$, or discards the message if exp > s). Otherwise, v accepts the message, assigns c the value 0, and assigns cmax a random integer in the range 0..N-1.

This modification achieves two objectives. First, it guarantees that process v never discards more than N data messages when the synchronization between u and v is lost due to some fault. Second, it ensures that the adversary cannot predict the instants when process v is willing to accept any received data message, and so cannot exploit such predictions by sending replayed data messages at those instants. Formally, process u and v in this protocol can be defined as follows. process u

```
var  nxt  : integer           {sequence number of}
                              {next sent message}
begin
     true →     send data(nxt) to v;   nxt := nxt + 1
end


process v
inp  N    : integer
var  s    : integer           {sequence number of}
                              {received message}
     exp  : integer           {sequence number of}
                              {next expected  message}
     c, cmax   : 0 .. N − 1
begin
     rcv data(s) from u →
         if s < exp ∧ c ≠ cmax →
              {reject message; report an adversary}
              c := (c + 1) mod N
         [] s ≥ exp ∨ c = cmax →
              {accept message}
              exp := s + 1;
              if c ≠ cmax →  c := (c + 1) mod N
              [] c = cmax →
                   c := 0;
                   cmax := RANDOM(0, N − 1)
              fi
         fi
end
```

Processes u and v of the soft sequence number protocol can be combined with process pw of the weak integrity protocol to construct process ps of the strong integrity protocol. A main difference between processes pw and ps is that pw exchanges messages of the form data(t, d), whereas ps exchanges messages of the form data(s, t, d), where s is the message sequence number computed according to the soft sequence number protocol, t is the message text, and d is the message digest computed over the concatenation (s; t; scr) of s, t, and the shared secret scr. Process ps in the strong integrity protocol can be defined as follows. (Process qs can be defined in the same way.) process ps

```
inp  sp         : integer
     sq         : array [0 .. 1] of integer
     N          : integer
var  s, t, d    : integer
     exp, nxt   : integer
     c, cmax    : 0 .. N − 1
begin
     rcv data(s, t, d) from qs →
         if MD(s; t; sq[0]) = d ∨ MD(s; t; sq[1]) = d →
              if s < exp ∧ c ≠ cmax →
                   {reject message and}
                   {report an adversary}
                   c := (c + 1) mod N
              [] s ≥ exp ∨ c = cmax →
                   {accept message}
                   exp := s + 1;
                   if c ≠ cmax →  c := (c + 1) mod N
                   [] c = cmax →
                        c := 0;
                        cmax := RANDOM(0, N − 1)
                   fi
              fi
         [] MD(s; t; sq[0]) ≠ d ∧ MD(s; t; sq[1]) ≠ d →
                   {report an adversary} skip
         fi
     []  true →
```

```
                    fi
[]    true →
         {p receives a data(s, t, d) from a router}
         {other than q and checks that its encryption}
         {is correct and its sequence number is}
         {within range}
         RTMSG
[]    true →
         {either p receives a data(t) from adjacent host}
         {or p generates the text t for the next data}
         {message}
         RTMSG

end
```

The first and second actions of process ps have a statement RTMSG that is defined as follows.

```
    if NXT(t) = p →          {accept message} skip
    [] NXT(t) = q →          d := MD(nxt; t; sp);
                             send data(nxt, t, d) to qs;
                             nxt := nxt + 1
    [] NXT(t) ≠ p ∧ NXT(t) ≠ q →
         {compute next soft sequence number s;}
         {compute d as the message digest of the}
         {concatenation of snxt, t and the secret}
         {for sending data to NXT(t); forward}
         {data(s, t, d) to router NXT(t)} skip
    fi
```

A proof of the correctness of the strong integrity protocol is presented in the full version of the paper [6].

IV.   IMPLEMENTATION CONSIDERATIONS

In this section, we discuss several issues concerning the implementation of hop integrity protocols presented in the last three sections. In particular, we discuss acceptable values for the inputs of each of these protocols. There are four inputs in the secret exchange protocol in Section 3. They are R p, B q, te and tr. Input R p is a private key for router p, and input B q is a public key for router q. These are long-term keys that remain fixed for long periods of time (say one to three months), and can be changed only off-line and only by the system administrators of the two routers. Thus, these keys should consist of a relatively large number of bytes, say 1024 bytes each. There are no special requirements for the encryption and decryption functions that use these keys in the secret exchange protocol. Input te is the time period between two successive secret exchanges between pe and qe. This time period should be small so that an adversary does not have enough time to deduce the secrets sp and sq used in computing the integrity checks of data messages. It should also be large so that the overhead that results from the secret exchanges is reduced. An acceptable value for te is around 4 hours. Input tr is the time-out period for resending a rqst message when the last rqst message or the corresponding rply message was lost. The value of tr should be an upper bound on the round-trip delay between the two adjacent routers. If the two routers are connected by a high speed Ethernet, then an acceptable value of tr is around 4 seconds. Next, we consider the two inputs sp and sq and function MD used in the integrity protocols in Sections 4 and 5. Inputs sp and sq are short-lived secrets that are updated every 4 hours. Thus, this key should consist of a relatively small number of bytes, say 8 bytes. Function MD is used to compute the digest of a data message. Function MD is computed in two steps as follows. First, the standard function MD5 [15] is used to compute a 16- byte digest of the data message. Second, the first 4 bytes from this digest constitute our computed message digest. As discussed in Section 5, input N needs to be much larger than 1. For example, N can be chosen 200. In this case, the maximum number of messages that can be discarded wrongly whenever synchronization between two adjacent routers is lost is 200, and the probability that an adversary who replays an old message will be detected is 99 percent. The message overhead of the strong integrity protocol is about 8 bytes per data message: 4 bytes for storing the message digest, and 4 bytes for storing the soft sequence number of the message.

V.   CONCLUDING REMARKS IN THIS PAPER,

i introduced the concept of hop integrity in computer networks. A network is said to provide hop integrity iff whenever a router p receives a message supposedly from an adjacent router q, router p can check whether the received message was indeed sent by q or was modified or replayed by an adversary that operates between p and q. I also presented three protocols that can be used to make any computer network provide hop integrity. These three protocols are a secret exchange protocol (in Section 3), a weak integrity protocol (in Section 4), and a strong integrity protocol (in Section 5). These three protocols have several novel features that make them correct and efficient. First, whenever the secret exchange protocol attempts to change a secret, it keeps both the old secret and the new secret until it is certain that the integrity check of any future message will not be computed using the old secret. Second, the integrity protocol computes a digest at every router along the message route so that the location of any occurrence of message modification can be determined. Third, the strong integrity protocol uses soft sequence numbers to make the protocol tolerate any loss of synchronization. All three protocols are stateless, require small overhead at each hop, and do not constrain the network protocol in any way. Thus, we believe that they are compatible with IP in the Internet, and it remains to estimate or measure the performance of IP when augmented with these protocols.

REFERENCES

[1]   TCP SYN Flooding and IP Spoofing Attacks", CERT Advisory CA-96.21, available at http://www.cert.org/.
[2]   S. Cheung, "An Efficient Message Authentication Scheme for Link State Routing", Proceedings of the 13th Annual Computer Security Applications Conference, San Diego, California, December 1997, pp. 90-98.
[3]   D. E. Comer, Internetworking with TCP/IP: Vol. I: Principles, Protocols, and Architecture, Prentice-Hall, Englewood Cliffs, NJ, 1988.
[4]   P. Ferguson and D. Senie, "Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address Spoofing", RFC 2267, January 1998. [5] M. G. Gouda, Elements of Network Protocol Design, John Wiley & Sons, New York, NY, 1998. [6] M. G. Gouda, E. N. Elnozahy, C.-T. Huang, T. M. McGuire, "Hop Integrity in Computer Networks", Technical Report TR-00-19, Department of Computer Sciences, The University of Texas at Austin, August 2000.
[5]   L. Joncheray, "A Simple Active Attack Against TCP", Proceedings of the 5th USENIX UNIX Security Symposium, 1995, pp. 7-19.
[6]   [8] S. Kent and R. Atkinson, "Security Architecture for the Internet Protocol", RFC 2401, November 1998.
[7]   S. Kent and R. Atkinson, "IP Authentication Header", RFC 2402, November 1998.
[8]   S. Kent and R. Atkinson, "IP Encapsulating Security Payload (ESP)", RFC 2406, November 1998. [11] S. Murphy and M. Badger, "Digital Signature Protection of the OSPF Routing Protocol", Proceedings of the 1996 Internet Society Symposium on Network and Distributed Systems Security, San Diego, California, February 1996. [12] D. Maughan, M. Schertler, M. Schneider, and J. Turner, "Internet Security Association and Key Management Protocol (ISAKMP)", RFC 2408, November 1998.
[9]   H. Orman, "The OAKLEY Key Determination Protocol", RFC 2412, November 1998.
[10]   J. Postel, "Internet Control Message Protocol", RFC 792, September 1981. [15] R. L. Rivest, "The MD5 Message-Digest Algorithm", RFC 1321, April 1992.
[11]   [W. R. Stevens, TCP/IP Illustrated, Vol. I: The Protocols, Prentice-Hall, Englewood Cliffs, NJ, 1994.
[12]   B. Smith, S. Murthy, and J. J. Garcia-Luna-Aceves, "Securing Distance Vector Routing Protocols", Proceedings of the 1997 Internet Society Symposium on Network and Distributed Systems Security, San Diego, California, February 1997.
[13]   M. de Vivo, G. de Vivo, and G. Isern, "Internet Security Attacks at the Basic Levels", Operating Systems Review, Vol. 32, No. 2, SIGOPS, ACM, April 1998.

BIOGRAPHY

Ashis Saklani :- Received M.tech in Computer Science from Karnataka University, and MCA from Graphic Era Institute of Technology (Now Graphic Era University) Dehradoon.He is a CISCO And Microsoft Certified Professional. He worked as a VPN Support Engineer at HCL Infinet, Noida.Presently he is working as Assistant Professor, at H.N.B Garhwal Central University, Srinagar, Uttarakhand, India.

Effect of Silica Fume on Engineering Properties of Black Cotton Soil

Chhaya Negi[1] , R.K.Yadav[2] , A.K. Singhai[3]

[1]*M.E. Scholars, Geotechnical Engineering, JEC Jabalpur,(M.P.),India ,*
[2] *Associate Professor, Civil Engineering Deptt., JEC Jabalpur,(M.P.),India ,* [3]*Associate Professor, Civil Engineering Deptt., JEC Jabalpur,(M.P.),India*

ABSTRACT:

Due to rapid growth of urbanization and industrialization, minimization of industrial waste is serious problem in present days. To encounter this ennovative and nontraditional research on waste utilization is gaining importances now a days. Soil improvement using the waste material like Slags,Rice husk ash,Silica fume etc.,in geotechnical engineering has been recommanded from environmental point of view. The main objective of this study is to evaluate the feasibility of using Silica fume as soil stabilization material. In this paper the effect of Silica fume on engineering characteristics of expansive clay like Black Cotton Soil has been presented. A series of laboratory experiment has been conducted on black cotton soil blended with Silica fume content from 5% to 20% by weight of dry soil. The experimental results showed a significant increase in California bearing ratio and Unconfined compressive strength.The Differential free swell of the clay is reduced from 50% to 7% with increase in Silica fume content from 0% to 20% respectively.The Proctor compaction results showed a small decrease in Maximum dry density and increase in Optimum moisture content. From this investigation it can be concluded that the Silica fume as a potential to improve the characteristics of black cotton soil.

KEYWORDS: *Silica Fume. Black cotton soil. Soil Stabilization. Engineering Properties.*

I.   INTRODUCTION:

Black cotton soil (BC Soil) represents a well known category of problematic from civil engineering point of view. They exibit large volumetric changes shrinkage and swelling behaviour if the moisture content changed.Due to this nature this type of soil is susceptible to damage to the structures and pavements founded on it.In india expansive soils cover about $0.8X10^6$ km² area approximately 20% of surface area[1].Structure founded in areas with soft or weak soil have need for improvement of soil properties by using additives.Soil stabilization techniques are used to improve shear strength,CBR, reducing expansive characteristics, etc.Silica fume also referred as micro-silica is a product resulting reduction of high purity quartz with coal in an electric arc furnace in the manufacture of silicon or ferro-silicon alloy.Silica fume rises as an oxidized vapour. It cools,condenses and is collected. It is fine grey coloured powder sometime similar to Portland cement or some flyashes.Condensed silica fume is essentially silicon-dioxide (more than 90%) in non crystalline form. Since it is an air borne material line flyash it has spherical shape.It is extremely fine with a particle size less than 0.1 micron and specific surface area of about 20,000m²/kg.Silica fume is used as an artificial pozzolanic admixure in concrete. As far as the production of silica fume is concerned nearerly 100,000 tons of micro silica is produced each year world wide[2].Iron also has a large amount of micro silica production. Steel Authority of India has provided necessary facilities to produced more than 3000 tons of Silica fume annually[3].

Many waste materials are used to modify the characteristics of soft soils. Traditionally the soils are stabilized by lime ,cement, etc.In recent year the uses of waste materials like flyash,plastic, rice- husk ash,slag,etc. for soil stabilization is gaining importance.In this study attempts are made to find the influences of silica fume on engineering characteristics of black cotton soil.

II.   LITERATURE REVIEW:

The engineering properties of clayey subgrade soils may need to be improved to make them suitable for construction using some sort of stabilization methods. Stabilization of pavement subgrade soils has traditionally relied on treatment with lime,cement, or waste materials such as flyash,slags,Silica Fume,etc.Many researchers

are looking for alternative materials for soil stabilization, fly ash is an effective agent for chemical and mechanical stabilization of soil.[4-6]

Saranjeet Rajesh Soni et. al. concluded that solid waste disposal is an economical and effective way to achieve improvement in engineering performance of black cotton soils. The stability of soil using fly ash and rice husk powder can be increased .[7] Abd.El-Aziz M.et al.(2004), examined the effect of lime-silica fume stabilizers on engineering properties of clayey subgrades. They summerised that the plasticity index and swell potential decreases and CBR value increases significantly. There is improvement in shear strength parameter also. [8] .Azzawi et al.(2012) studied effect of silica fume addition on behaviour of silty clayey soils, they investigated that there is significant important on swelling pressure and compressive strength of composite samples with silica fume.The permeability of soil increased with increase in silica fume content. It is observed that the addition of silica fume decreases the development of cracks on the surface of compacted clay samples reducing the cracks width by 75%.[9]Venu Gopal N., studied the soil properties with silica fume as stabilizer and comparing the same with other materials. The laboratory investigations indicate that soil samples possessing low strength can be treated with varying silica fume of 5% to 20% by weight of dry soil. The treated soil samples showed significant improvement in the strength characteristics.[10]

Biswas et al.(2012) studied the utilization of rice husk with lime in subgrade soil for a rural road;they concluded that a very little amount of lime (3%) added to the clayey soil with RHA, improve the CBR value and compaction characteristics to a great extent.[11]The effect of marble dust with RHA on expansive soil has been studied by Sabat and Nanda (2011) it has been reported that the CBR and UCS values increase substantially due to addition of these two materials with natural expansive soil.[12] Kalkan and Addulut (2004) examined the suitability of silica fume for the construction of hydraulic barrier in landfill. The concluded that clay mixed with silica fume in different proportions, has higher binding strength.,low swelling pressure, and high compressive and shear strength.[13] M.Karimi and A.Ghorbani (2011) studied effect of lime and microsilica admixtures on silty sand soils, in presence of sulphates.Results showed that the addition of microsilica to the silty sand soil increases the CBR strength and decreases swelling, therefore microsilica waste material can be successively used to enhance the strength of silty soil.[2]

III.   MATERIALS AND METHODOLOGY:

In soil, Silica fume is mixed on percentage basis i.e. 0%, 5%, 10%,15% and 20% by weight of dry soil.The following Tests were conducted on BC soil and Silica Fume mixes ;as per relevant IS Code.
- Compaction Test
- California Bearing Ratio
- Unconfined Compressive Strength
- Differential Free Swell

SILICA FUME

The stabilizer materials used in this study was Silica Fume . Silica Fume used in this study brought from Shisher Export House, Raipur,(C.G.). The composition of SF is presented in table-1.

CHEMICAL PROPERTIES

Table-1,Chemical composition of silica fume

| S.No.	Parameters	Test Value
I	Chemical Test	
1.	Silica as SiO_2, % by mass	89.9
2.	Total Sulphur Content as SO_3, % by mass	0.58
3.	Lime as CaO, % by mass	7.85
4.	Magnesia as MgO, % by mass	4.03
5.	Alumina as Al_2O_3, % by mass	Nil
6.	Iron Oxide as Fe_2O_3, % by mass	Nil
II	Physical Test	
1.	Density , g/cc	2.07
2.	Particle Size Distribution (%)	59

Black Cotton Soil

   The black cotton soil has been used as a base material in this study.It has been replaced partially by silica fume by weight of dry soil.The clayey soil involved in this research was collected from Railway Station Nursery Area, Adhartal, Jabalpur (M.P.). The soil sample was disturbed. The soil is classified as clay of high plasticity (Gs = 2.71 with 95% fines) with expansive behavior. The engineering characterictics of clay sample are presented in Table-2.

Table-2,Engineering Properties of black cotton soil

Sr.No.	Particulars	Test results
1	Soil Classification	CH
2	Specific Gravity	2.71
3	Liquid Limit , (%)	53
4	Plastic Limit , (%)	27.52
5	Plasticity Index , (%)	25.48
6	Shrinkage Limit , (%)	7.55
7	Grain Size Distribution: Sand (%) Silt + Clay (%)	5 95
8	Compaction Characteristics Optimum Moisture Content (%) Maximum Dry Density (kN/m³)	17 1.58
9	California Bearing Ratio	1.63
10	Swelling Characteristics Differential Free Swell (%)	50
11	Unconfined Compressive Strength (kN/m²)	qu=125.1

The grain size distribution curve of BC soil is shown in figure 1
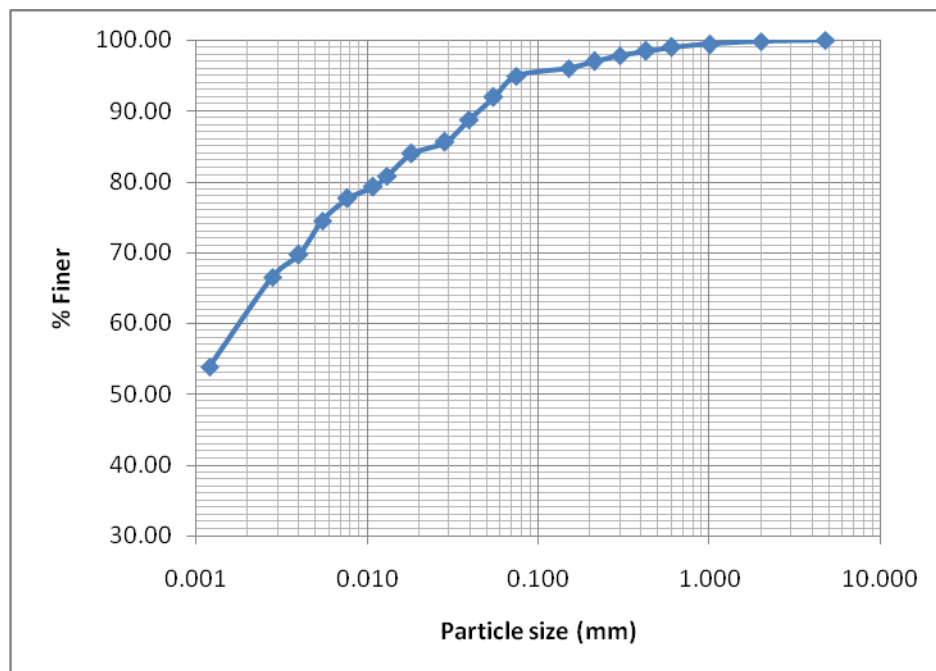
Figure-1, Grain Size Distribution of BC soil

IV. RESULTS OF VARIOUS TESTS

The results of various tests are summarized in table -3

Table-3, Results of various tests

Sr.No.	Particulars	Observation & Results				
		CS0	CS5	CS10	CS15	CS20
I	Compaction Characteristics Optimum Moisture Content (%) Maximum Dry Density (kN/m²)	17 1.58	18 1.55	23 1.53	27 1.51	28 1.50
II	California Bearing Ratio	1.63	2.37	2.51	2.60	2.81
III	Swelling Characteristics Differential Free Swell (%)	50	25	15	11	7
IV	Unconfined Compressive Strength (kN/m²)	qu=125.1	qu=127.6	qu=136.0	qu=155.6	qu=163.6

Where, CS0 = Clay + 0% Silica Fume
CS5 = Clay + 5% Silica Fume
CS10 = Clay + 10% Silica Fume
CS15 = Clay + 15% Silica Fume
CS20 = Clay + 20% Silica Fume

Result and Discussions:
The test results obtained from the experimental work are briefly discussed below.

The compaction tests have been carried out on virgin soil and soil-silica fume mixtures. The proctor tests were conducted as per IS 2720 (Part-VIII). The variation of optimum moisture content (OMC) and maximum dry density (MDD) are shown in figure 2. Both the OMC and MDD decrease with increase in silica fume content.However this decrease is not much as compare to the initial values. Soaked CBR tests were performed to evaluate the strength properties of the soil and mix.The CBR values increase from 1.63% to 2.81% as the silica fume content increases from 0% to 20%. The variation of CBR values with increase in silica fume content is shown in figure 3. The swelling behaviour of the soil is also checked to a great extent. The differential free swell (DFS) values are decrease from 50% to 7%. The variation is shown in figure 4. There is significant improvement in Unconfined compressive strength of the black cotton soil. The Unconfined compressive strength increases from 125.1 kN/m² to 163.6 kN/m² with the increase in silica fume content 0% to 20%. The variation is shown in figure 5.
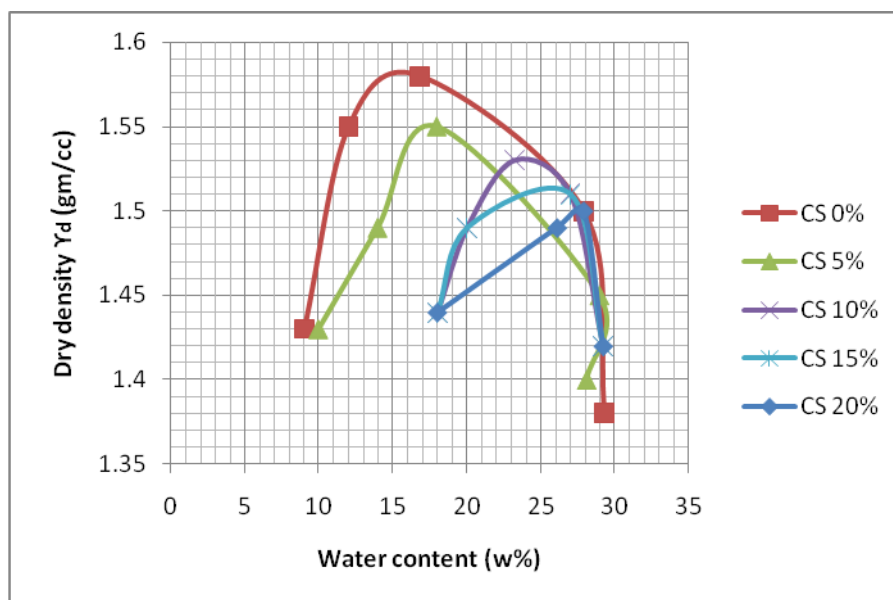
Figure 2 :Effect of Silica Fume on Compaction Characteristics of Soil.

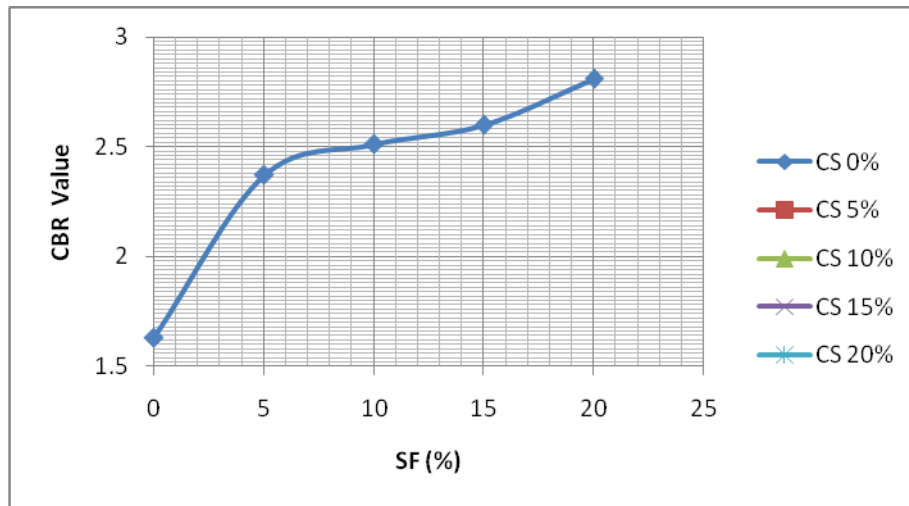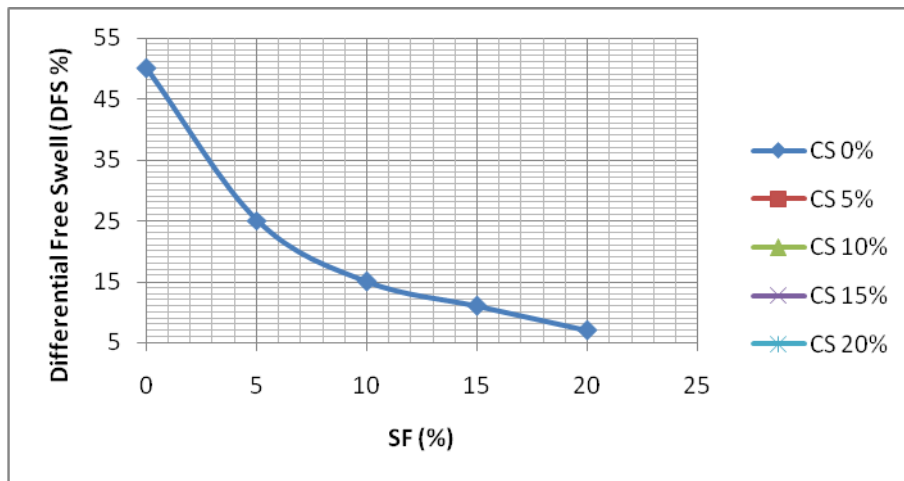**Figure 3: Effect of Silica Fume Content on California Bearing Ratio.**

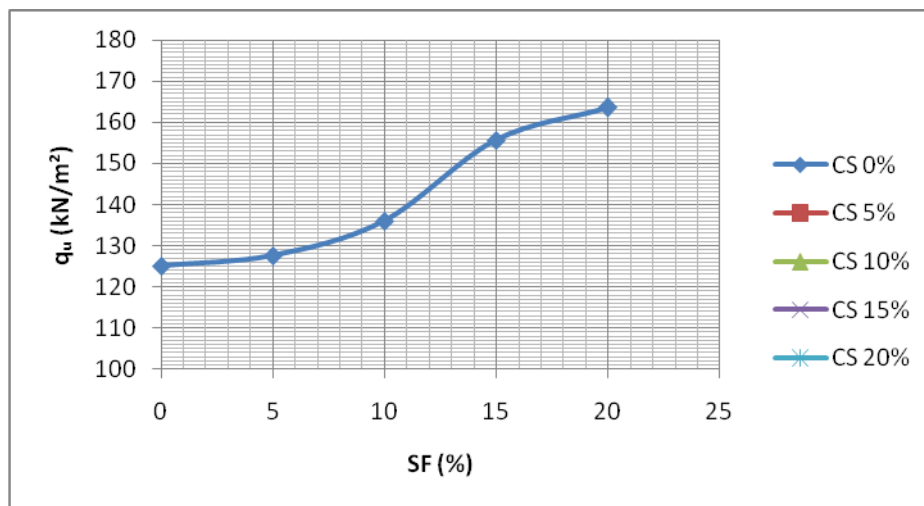**Figure 4: Variation of Differential Free Swell(DFS) with silica fume content.**

Figure 5: Variation of Unconfined Compressive Strength (qu) with silica fume content

V. CONCLUSIONS:

From the above laboratory investigations the following conclusions can be drawn :

[1]   The BC soil-Micro Silica change the proctor compaction parameters.The addition of silica fume to the black cotton soil increase the optimum moisture content and decreases the maximum dry density with the increase in silica fume content.

[2]   The addition of silica fume to the black cotton soil improve the soaked CBR considerably.The addition of 20% silica fume to the black cotton soil increases the CBR strength by 72% approximately.

[3]   There is a significant decrease in the swelling characterristics of the soil. The degree of expansiveness reduceses from "High to Low".

[4]   The addition of silica fume also increases the Unconfined compressive strength.The UCS of stabilized samples significantly increases from 125.1 kN/m² to 163.6 kN/m² i.e. approximately 31% increase.

[5]   From the laboratory investigation it can be concluded that the industrial wastes like silica fume has the potential to modify the engineering characteristics of expansive clay like black cotton soil.

REFERENCES:

[1]   Choudhary AK, Gill KS & Jha KN (2011), "Improvement in CBR values of expansive soil subgrade using geosynthetics,"Proc. Indian Geotechnical Conference, 569-572.

[2]   M Karimi & Ali G."Stabilization of silty sand soils with lime and micro silica admixture in pressure of sulphates".Pan –Am CGS ,Geotechnical Conference 2011.

[3]   Shetty MS 'Concrete Technology' 2012, S.Chand and Co. Ltd. New Delhi.

[4]   "Fly Ash for Soil Improvement", Geotechnical Special Publication No.36, America Society of Civil Engineers, New York,1993.

[5]   Sherwood, P.T. (1995). "Alternative materials in road construction".Thomas Telford Publications, London, 1995.

[6]   "Soil and Pavement Base Stabilization with Self Cementing Coal Fly Ash", American Coal Ash Association, Alexandria, Virginia, May 1999.

[7]   Saranjeet Rajesh Soni *et al. "Disposal of solid waste for black cotton soil stabilization",International Journal of Advanced Engineering Sciences and Technologies Vol No. 8,Issue No. 1,113 -120

[8]   Abd El-Aziz.,Abo-Hashema M., and El-Shourbagy M., "The effect of Lime-Silica Fume Stabilizer on Engineering Properties of Clayey Subgrade",.Fourth Mansoura International Engineering Conference (4th IEC), Faculty of Engineering ,Mansoura University,Eygpt,April 2004.

[9]   Dr. Adel A. Al-Azzawi, Khalida A. Daud ,Muhammed A. Abdul Sattar "Effect of Silica Fume Addition on the Behavior of Silty-Clayey Soils" Journal of Engineering and Development, Vol. 16, No.1, March 2012.

[10]   Venu Gopal.N,Project report on "Study of Soil Properties with Silica Fume as Stabilizer and Comparing the same with RBI-81 and Cost Estimation",PG diploma in Highway Engineering,Visvesvaraya Technological University,Belgaum.

[11]   Biswas, etal. (2012) ."Utilization of Rice husk with Lime in Subgrade Soil for a Rural Road."International Conference on Emerging Frontier in Technology for Rural Area.

[12]   Sabat, A.K., Nanda,R.P. (2011). "Effect of marble dust on strength and durability of Rice husk ash stabilized expansive soil."International Journal of Civil and Structural Engineering,Volume 1,No 4, pp.-939-948.

[13]   Kalkan,E. and Akbulut,S., "The Positive Effects of Silica Fume on the Permeability,Swelling Pressure and Compressive Strength of Natural Clay Liners",Journal of Engineering Geology,Vol. 73,2004,pp. 145-156.

[14]   Muntohar, A.S., 1999a, "Effect of Lime and Rice Husk Ash on the Behavior of Soft Clay", Regional Seminar at Islamic University of Indonesia Sept. 4th 1999. Yogyakarta, Indonesia.

[15]   ACI Committee 226. 1987b. "Silica fume in concrete": Preliminary report. ACI Materials Journal March-April: 158-66.

[16]   Sivapullaiah, P.V., Sridharan, A., Bhaskar, R.K.V., " Role of Amount and Type of Clay in the Lime Stabilization of Soils", Ground Improvement Vol. 4, 2000, pp. 37– 45.

[17]   Chen, F.H., 1981. Foundation on Expansive Soil, Development in Geotechnical Engineering 12. Elsevier Scientific Publishing Company, New York, USA.

[18]   Bell, F.G., "Engineering Treatment of Soils", Chapmen and Hall, Publishers, London, 1993.

[19]   Yoder, E.J. and Witczak, M.W., 1975. Principles of Pavement Design, 2nd Edition, John Wiley & Sons, Inc., USA.

[20]   Silica Fume User's Manual;US Department of Transportation,Federal Highway Administration,FHWA-IF-05-016,April 2005.

Design Of A Low Power And High Speed 1.5 Bit Stage For Pipeline ADC

[1,]Darshana Upadhyay, [2,] Sudha Nair[1,]

[1,2,]*Asst. Professor, Department of Electronics & Communication, R.K.D.F.I.S.T, Bhopal, M.P.)*

ABSTRACT

The design and the preliminary measurements of a prototype 10 bit pipeline ADC based on 1.5-bit per stage architecture is presented here. For conventional ADCs differential amplifiers dominate the power dissipationin most high-speed pipelined analog-to digital converters (ADCs).We propose a new technique to design Pipeline ADCwith Gm stage as differential amplifier for low power consumption. The gm based amplifier performs a class-AB operation by smoothly changing between a comparator-like semi digital driver and a continuous– time high-gain amplifier according to the input voltage difference.The ADC has been designed with 50nm CMOS technology and achieves the low power upto 1.5 mWand integral nonlinearity within 0.5 LSB and 0.9 LSB, respectively.

Keywords: Gm-based amplifier, operational transconductance amplifier (OTA), Op-Amp, Pipeline Analog-to-digital converter (ADC),

I.   INTRODUCTION

PIPELINED analog-to-digital converters (ADCs) have lowinput capacitance, high-speed concurrent operation, andhardware complexity that linearly increase with resolution.Accordingly, among various ADC structures, the pipelinedADC is a popular architecture, because it achieves high speedand resolution while maintaining area efficiency and low powerconsumption [6]. As a result, it is a suitable architecture for implementingADCs for video and communication applications.The conventional ADC requires Op-Amp with high power consumption.µcrossing-based circuits [1], dynamic source-followeramplification [2], capacitive charge pump [3], digitally assistedalgorithmic amplifier [4], and ring amplifier [5].These techniques require the additional calibration circuits to reduce thedependency on the stability on parametervariations. The amplifier less technique available in [7] presents a Gm-based amplifier that follows a conventional two-phaseswitched-capacitor circuit. Without any calibration techniqueor digital control, the proposed amplifier achieves a smoothtransition from a comparator-like semidigital operation to acontinuous-time high-gain amplifier whenever needed.

In this paper, we describe the implementation and measurement results of a 10-bit 20 MSamples/s pipelinedADC that uses Gm based amplifier.In order to facilitate and expedite the evaluation of the proposedapproach, we based our design on an existing commercially available pipelinedADC in 0.5µmCMOStechnology [8]. In Section II, we briefly review the architecture of this pipeline ADC. Section III highlights the architecture of Gm based amplifier and section IV gives design of 1.5 bit stage for pipeline ADC architecture. In Section I,we establish the simulation and measuredresults. Finally section VI gives the conclusion and future scope.

II.   PIPLINE ADC DESIGN

The pipeline ADC is built of a number of serially connectedconverting stages as shown in Fig. 1. In this work architecturewith 1.5-bit stages is chosen because of its simplicity and immunityto the offsets in the comparator and amplifier circuits [6].The 1.5-bit since generates only three different values coded on2 output bits which are sent to a digital correction block where18 input bits from 9 stages are combined together resulting in10 bits of ADC output.The block diagram of fully differential single stage is shown in Fig. 2. Each 1.5-bit stage consist of two comparators, two pairs of capacitors Cs and Cf , a Gm amplifier, several switches and small digital logic circuit. The stage gain of 2 is obtained setting Cs = Cf .
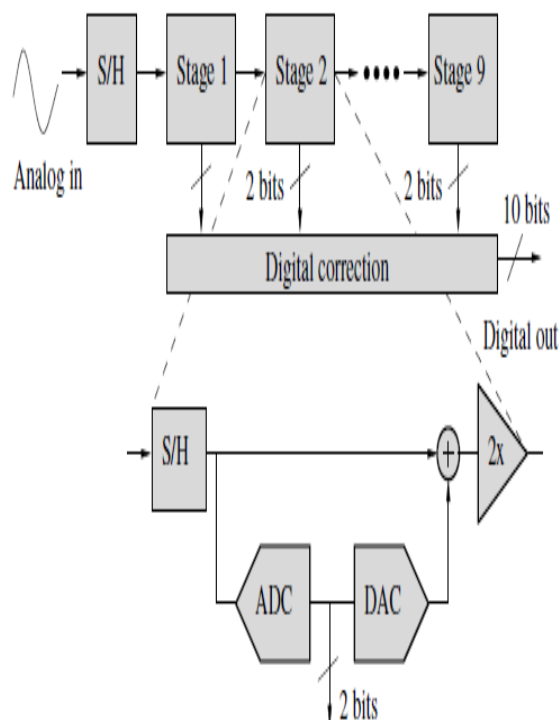
Fig.1 Pipeline ADC architecture

Since the chosen ADC architecture leaves   very relaxed requirements on the comparators thresholds (100 mV precision) the comparators are designed as simple dynamic latches. The power consumption block of this ADC is the Gm stage, so a further improved design for this stage is also presented here.
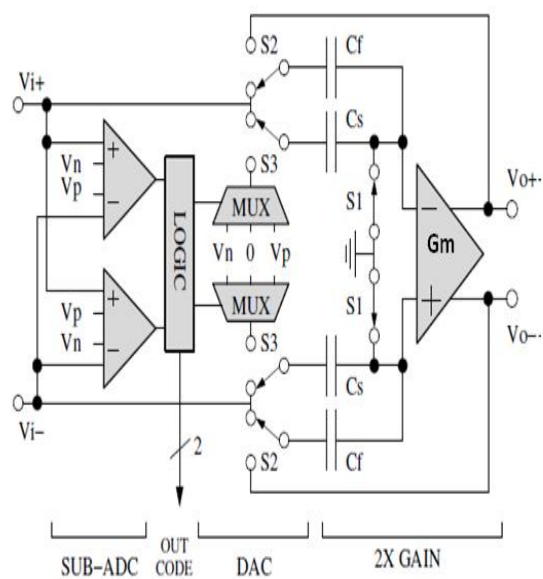
Fig.2 Simplified schematic of a 1.5-bit stage

III.   GM-BASED AMPLIFIER
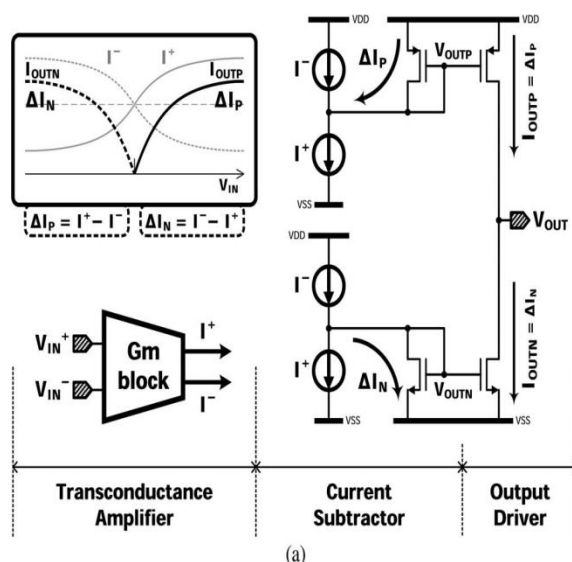
Fig. 3 Concept of the proposed Gm-based amplifier

A new technique to design an Gm based amplifier to reduce the power consumption is proposed in [8]. Fig.3 shows the concept of the proposed Gm-based amplifier, which have a Gm block, a current subtractor stage, and an output driver stage. The amplifier gain is the product of transconductanceGm and the output resistance of the output stage. While Gm does not change much by bias control, the output resistance of a transistor significantly increases as the gate-to-source bias decreases. If the transistors in the output driver are biased to operate in subthreshold region, which almost turns the transistors off, the output resistance approaches to infinity, resulting in an infinite gain of the amplifier.Fig 4 shows the complete circuit structure for Gm based amplifier. The biasing of the output transistors in deep-subthreshold region gives additional benefits. Since the resistance of the output transistors and the load capacitance are much larger than the resistance of the current sources and the parasitic capacitance in the current subtractor, there is no need for an additional frequency compensation scheme. In addition, the proposed amplifier does not require common-mode feedback when used in switched-capacitor circuits driven by nonoverlap-ping two-phase clocks .

Fig.4 Gm-based amplifier

IV.   DESIGN OF 1.5 BIT STAGE FOR ADC

In The complete design of 1.5 bit stage for Pipeline ADC is shown in figure 5. This stage consists the Gm based amplifier, a dynamic comparator and a switched capacitor network with some digital switches. Both the capacitor has the sa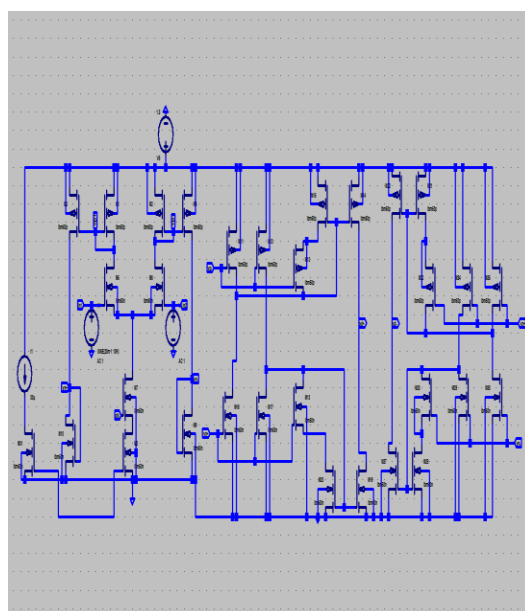me values.During the sampling phase CP1, the comparator produces a digital output Di. Di is 1 if Vin>Vthand Di is 0 if Vin <Vth, where Vth is the threshold voltage defined midway betweenVref+ and Vref-. During multiplying phase CP2, Cs is connected to the output of the operational amplifier and Cf is connected to either the reference voltage Vref+ or Vref-, depending on the bitvalue Di.

If Di = 1, Cf is connected to Vref+, resulting in the residue(Vout) is :

Vout (i) = 2 Vin(i) –Di.Vref+………(1)
Otherwise, Cf is connected to Vref+, giving an output voltage:
Vout (i) = 2 Vin(i) - Di.Vref-………..(2)

The comparators are the simple dynamic comparators because of less requirements of threshold voltages.
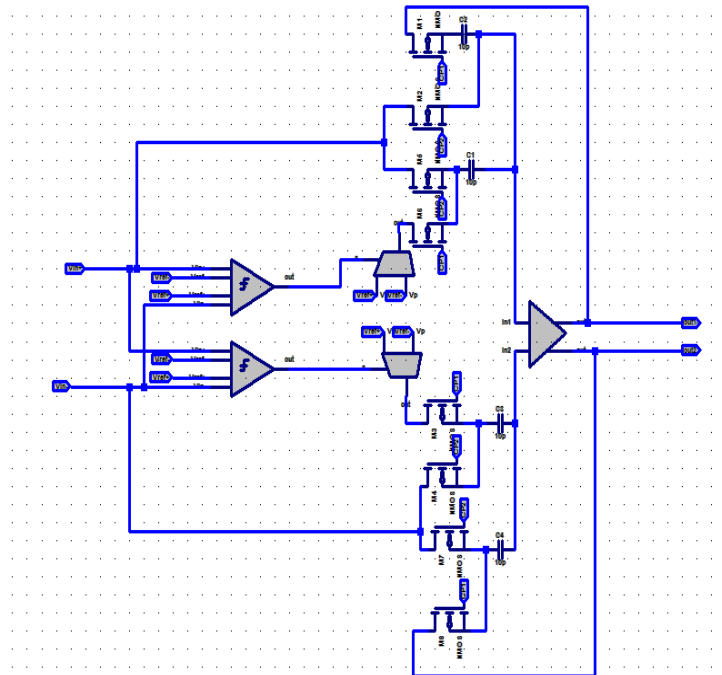
Fig 5 1.5 bit stage design

The 1.5-bit DAC is basically an analog multiplexer controlled by the flash ADC's output. It must be noted that, in the fully differential circuit, the voltage *Vref* is obtained by swapping the two *Vref* lines and, therefore it is accurate giving a linear DAC. The SHA provides the analog memory needed for pipelining and does some analog arithmetic. Its output voltage, assuming ideal components, is:

$$V_{res} \begin{cases} 2\,Vin + Vref\,for\,Vin < \dfrac{+Vref}{4} \\ 2Vin\,for - \dfrac{Vref}{4} < Vin < \dfrac{+Vref}{4} \\ 2Vin - Vref\,for\,Vin > \dfrac{+Vref}{4} \end{cases}$$

The input gain of the SHA must be 2 for a proper ADCoperation, but this only happens if the two capacitors *Cs* are perfectly matched and the open-loop gain of the opampis infinite. In our design the sampling capacitance is limited by matching rather than noise because we still have a high voltage swing.

V. SIMULATION RESULTS.

The complete architecture was designed using 50 nm CMOS technology. Fig. 6 shows the frequency response of Gm based amplifier with the Dc gain of 70 dB with unity gain bandwidth of 200MHz. The power delivered by this amplifier is 0.24mW.
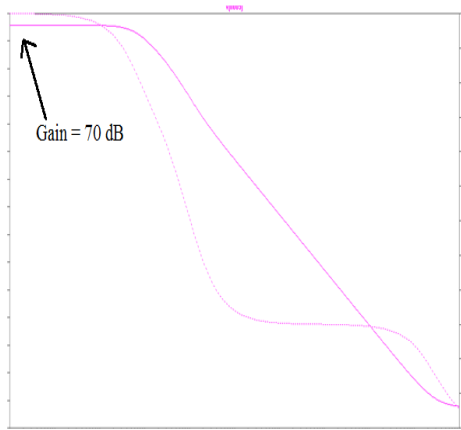
Fig.6 Gain plot for Gm based amplifier

The ADC accepts a full-scale differential input signal of2V peak to peak and has a differential capacitive loading of 3pF. The DNL and INL plots are shown in fig. 7 The measured integral non linearity is 0.6 LSB.The SFDR at 50MHz input is 100dB for the particular stage.The FFT plot for 50MHz samples per second is shown in the fig.8.The simulated results shows that the power dissipated by complete 1.5 bit stage is 1.5 mWat the full conversion speed of 50 MS/s. Table Isummarizes the measured specifications of the prototype ADCs
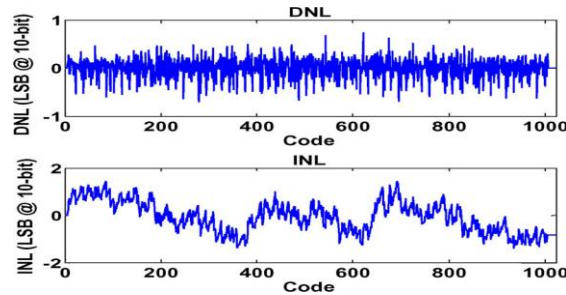
Fig.7 DNL and INL plot for 1.5 bit stage
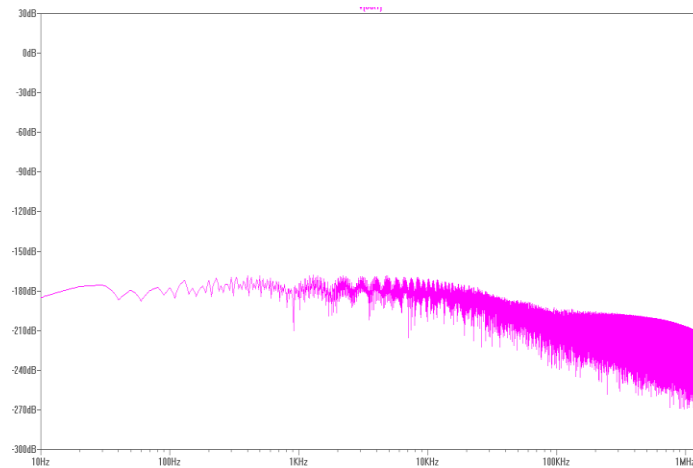
Fig.8 The FFT plot for input frequency of   50MHz

Table I ADC Performance

Parameter	Value
Conversion rate	50 Ms/s
Input range	2 V
Amplifier gain	70 dB
Power supply	1.8V
Power consumption	1.5mw
SFDR (50MHz input)	100dB

VI.   CONCLUSION

This paper proposes a new pipeline ADC architecture, which takes advantage of a large stage resolution in the first stage of the pipeline. This architecture shows the low power consumption in the pipeline concept. The prototype ADC demonstrates the ability of this new architecture.The architecture achieves a calibration-free, highresolution,moderate-speed, area-efficient and power-efficientADC design, which is difficult to achieve with traditional ADCs.
.

REFERENCE

[1]   L. Brooks and H.-S. Lee, "A 12b, 50 MS/s, fully differential zero-crossingbased pipelined ADC," IEEE J. Solid-State Circuits, vol. 44, no. 12,pp. 3329–3343, Dec. 2009.

[2]   J. Hu, N. Dolev, and B. Murmann, "A 9.4-bit, 50-MS/s, 1.44-mWpipelined ADC using dynamic source follower residue amplification,"IEEE J. Solid-State Circuits, vol. 44, no. 4, pp. 1057–1066, Apr. 2009.

[3]   I. Ahmed, J. Mulder, and D. A. Johns, "A low-power capacitive chargepump based pipelined ADC," IEEE J. Solid-State Circuit, vol. 45, no. 5,pp. 1016–1027, May 2010.

[4]   D.-W. Jee, S.-J. Park, H.-J.Park, and J.-Y. Sim, "A low-voltage OPamp with digitally controlled algorithmic approximation," in Proc. IEEECICC, 2008, pp. 499–502.

[5]   B. Hershberg, S. Weaver, K. Sobue, S. Takeuchi, K. Hamashita, andU. Moon, "A 61.5dB SNDR pipelined ADC using simple highly-scalablering amplifiers," in Proc. Symp. VLSI Circuits, 2012, pp. 32–33.

[6]   F. Maloberti, F. Francesconi, et al., Design considerationson low-voltage low-power data converters. IEEE Trans.Circuits Syst.I, 42, 853–863, 1995.

[7]   YunjaeSuh, Jongmi Lee, Byungsub Kim, Hong-June Park and Jae-Yoon Sim, "A 10-bit 25-MS/s 1.25-mW Pipelined ADC With a SemidigitalGm-Based Amplifier"inIEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS,2013 .

[8]   T. B. Cho, P. Gray, A 10 b, 20 Msample/s, 35 mW pipelineA/D converter. IEEE J. Solid-State Circuits, 30, 166–172,1995.

Incorporation of Dstatcom in Radial Distribution Systems

[1,]K. Nirmala, [2,]N. Poorna Chandra Rao

[1,]*PG Student, Dept.of EEE SVPCET, Puttur, Chittoor Dist., A.P., India,*
[2,]*Associate Professor, Dept.of EEE SVPCET, Puttur, Chittoor Dist., A.P., India*

ABSTRACT

In this paper presents a method to determine the weakest bus in a radial distribution system. A unique& novel voltagestability indicator (VSI) can identify the condition of load buses with voltage collapse point of view.The voltage stability indicator is derived from the voltage equation of radial distribution system. The weakest bus voltage profile is improved by placing a DSTATCOM. The DSTATCOM is modeled to supply the required reactive power for compensation and to maintain the voltage magnitude of the node where DSTATCOM is placed as 1 p.u. The validity of the proposed VSI and DSTATCOM modeling is examined by a standard 33 bus radial distribution system. The results validate the proposed VSI and DSTATCOM models in large distribution systems.
.

KEYWORDS: *Distribution system,DSTATCOM,load flow, voltage stability, voltage stability indicator (VSI), Load modelling.*

I.   NOMENCLATURE

P_i, Q_i Active and Reactive sending end power respectively
P_{i+1}, Q_{i+1}Active and Reactive receiving end power
V_i, V_{i+1} Sending end and Receiving end voltage
L_dProposed voltage stability indicator
L_e Existing indicator
P_L, Q_L Active and Reactive power losses respectively

II.   INTRODUCTION

In recent days, there is a growing interest in operation and control of distribution networks. The most of the distributionsystems are radial in nature. The Radial distribution networks [2,4,8, 9] having a combination of industrial, commercial,residential and lightning loads are generally weak in nature because of high resistance to reactance ratio.The Voltage stability is an important factor to be considered in power system operation and planning since voltage instability would lead to system collapse. The problem of voltage stability [4] has been defined as inability of the power system to provide the reactive power or non-uniform consumption of reactive power bythe system itself. So, the voltage stability is a major concern in planning and assessment of security of large power systems in contingency situation, especially in developing countries because of non -uniform growth of load demand in the reactive power management side [3]. The loads generally play a key role in voltage stability analysis. So, the voltage stability is known as load stability.

The radial distribution systems have high resistance to reactance ratio, which may cause high power loss in radial system. In power system, the radial distribution system is one which may suffer from voltage instability.In this paper a new technique used for determination of voltage stability at load bus. By using this Voltage Stability indicator, the buses of the system which are weak in nature can be identified easily. So far much attention was paid on voltage stability analysis of transmission lines, researchers have paid very little attention to   develop a voltage stability indicator for radial distribution systems [1,2,6]. In case of the radial distribution system,providing demanding power to the entire load while maintaining voltage magnitude at an acceptable range is one of the major system constraints. There are two principal conventional means of maintaining voltages at an acceptable range in distribution systems are series voltage regulators and shunt capacitors. Conventional series voltage regulators are commonly used for voltage regulation in distribution system [7,8,9]. But these devices cannot generate reactive power and by its operation only force the source togenerate requested reactive power and they have quite slow response as these operations are step by step [10].Shunt capacitors can supply reactive power to the system but they are not capable to generate continuous variable reactive power. Another difficulty associated with the application of distribution capacitors is the natural oscillatory behavior of capacitors when they are used

in the same circuit with inductive components. Which results in the well-known phenomena of ferroresonance and/or self-excitation of induction machinery [10].

The FACTS devices concept was originally developed for transmission systems, but the similar idea has been started to be applied to distribution systems. The Distribution STATCOM (D-STATCOM) is a shunt connected voltage source converter which has been utilized to compensate power quality problems such as unbalanced load, voltage sag, voltage fluctuation and voltage unbalance. D-STATCOM is also utilized for the improvement of another aspect of power quality, i.e. voltage compensation in long term. In section 2 presents the derivation of Voltage stability indicator, load flow of radial distribution system and mathematical modeling of DSTATCOM and in the section3 the VSI and DSTATCOM are tested on a 33 bus radial distribution system and results are analysed. Section 4 summarizes the load modelling process of different- types load model in distribution systems Section 5 summarizes the main points and results of the paper.

III. BASIC THEORY

3.1 Mathematical derivation of voltage-stability indicator

The VSI is derived from voltage equation of the radial distribution system. The proposed indicator is given below considering a line of impedance R+jX is connected between two nodes as shown in the following Fig.1 where i and j are respectively two nodes of the branch and node I is sending end node[V(i)] and node j is receiving end node[V(j)]. Therefore, power flow direction is from node i to node j. The load at node j is{P(j)+jQ(j)}.The impedance of the branch is R(i)+jQ(i) if line shunt admittances are neglected, the current flowing through the line is given by:
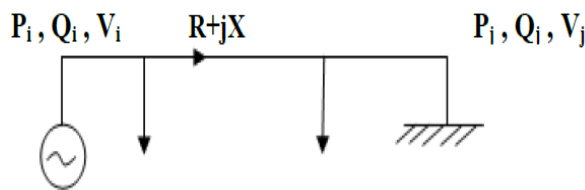
Figure 1: Radial system

$$I(i) = \frac{V_i \angle \theta_i - V_j \angle \theta_j}{R_i + j X_i} \qquad (1)$$

The complex power is given as: S=VI^*

$$I(i) = \frac{P(j) - j Q(j)}{V^*(j)} \qquad (2)$$

Equating (1) and (2)

$$\frac{V_i \angle \theta_i - V_j \angle \theta_j}{R_i + j X_i} = \frac{P(j) - j Q(j)}{V_j \angle - \theta_j}$$

$[V(i)\angle\theta_i - V(j)\angle\theta_j][V(j)\angle-\theta_j] = [P(j)-jQ(j)][R(i)+jX(i)] \qquad (3)$

Equating real& imaginary parts of (3) we get

$V(i)*V(j)\cos(\theta_i - \theta_j) - V(j)^2 = P(j)*R(i) + Q(j)*X(i) \qquad (4)$

$X(i)*P(j) - R(i)*Q(j) = V(i)*V(j)\sin(\theta_i - \theta_j) \qquad (5)$

As in radial distribution systems voltage angles are negligible, hence $(\theta_i - \theta_j) \approx 0$, equation (4),(5) becomes

$V(i)*V(j) - V(j)^2 = P(j)*R(i) + Q(j)*X(i) \qquad (6)$

$$X(i) = \frac{R(i) * Q(j)}{P(j)} \qquad (7)$$

From equation (6),(7)

$$V(j)^2 - V(i)*V(j) + P(j)R(j) + \frac{R(i)*Q(j)2}{P(j)} = 0 \qquad (8)$$

$$V(j)^2 - V(i)*V(j) + \frac{R(i)[P(j)2 + Q(j)2]}{P(j)} = 0 \qquad (9)$$

The roots of equation (6) are real if

$$V(i)^2 - \frac{4R(i)[P(j)2 + Q(j)2]}{P(j)} \geq 0$$

$$\frac{4R(i)[P(j)2+Q(j)2]}{V(i)2\,P(j)} \leq 0 \qquad (10)$$

$$VSI = \frac{4R(i)[P(j)2+Q(j)2]}{P(j)} \qquad (11)$$

Hence equation (11) is termed as voltage stability indicator (VSI).For stability of particular node the value of VSI must be VSI≤1. The range of VSI values is 0<VSI≤1. If the value of VSI approaches or greater than unity, then that node is highly unstable.

3.2 Load flow technique

In this paper, backward forward sweep load flow technique is used to compute voltages and power flow for a radial distribution system. Several methods have been developed based on the concept of doing backward forward sweeps of radial distribution networks [11,12,13].

Forward backward sweep based power flow algorithms generally take advantage of the radial network topology and consist of forward or backward sweep processes. In these algorithms, forward sweep is mainly the node voltage calculation from sending end to the far end of the feeder and laterals, and the backward sweep is primarily the branch current or power summation from far end to sending end of the feeder and laterals.Thefirst step to perform load flow is to create a matrix M (with columns as nodes and nodes as branches). For a particular branch (row) the sending end node of a branch is assigned -1 and receiving end as +1, remaining elements as zero. From that M matrix the column which doesn't have -1 represents end node. After identifying all end nodes, the back propagation path from end node to source node must be identified. After finding end node corresponding row gives the branch attached to it and in this row, -1 value is identified. The corresponding column gives the sending end node to the studied branch. This process continues until the algorithm reaches a column which has no element equal to 1 this represents source node. Hence the end nodes and back propagation paths for end nodes are calculated then employ KVL and KCL to radial distribution system, the power flow algorithms calculate node voltages in forward and backward processes. The radial part is solved by a straight forward two-step procedure in which branch currents are first computed (backward sweep) and then bus voltages are updated (forward sweep).

The voltage at node i can be expressed as

$$V(i) = V(i-1)-I(i)Z(i) \qquad (12)$$

The current at node i can be expressed as

$$I_L = \frac{PL(i)-QL(i)2}{V*(i)} \qquad (13)$$

Branch current can be given as

$$I(i) = I_i(i)+\sum_{j\in\beta_i} I(j) \qquad (14)$$

Algorithm for radial distribution system load flow:

[1] Read the distribution system line data& load data.
[2] From the node & branch matrix M.
[3] Get the end nodes& back propagation paths.
[4] Obtain the value of β of equation (14) by calculating the downstream nodes of every node.
[5] Make a flat start by assuming the voltage profile of all buses to be 1 p.u.
[6] Calculate load current IL(i) of each bus using equation (13)
[7] Summation of all the load currents corresponding to the nodes which are downstream to the desired node, as well as its own node; gives the current injected I(i) at that node.
[8] After calculating the current injected to each bus, calculate the voltage of each bus using equation (12).
[9] Compare the difference between each consecutive voltages values of every node. This gives deviation.
[10] If deviation is less than or equal to tolerance limit,then update new voltage values and go to step 6.Else display absolute value of voltage and the phase angle.
[11] Stop.

3.3 Mathematical modelling of DSTATCOM

The D-STATCOM is used for voltage regulation in the steady-state condition and can inject only reactive power to the system. Consequently, I_{dstat} must be kept in quadrature with voltage of the system. By installing DSTATCOM in distribution system, all nodes voltages, especially the neighboring nodes of

D-STATCOM location, and branch current of the network, change the steady state condition. The diagram of buses *i* and *j* of the distribution systems, when D-STATCOM is installed for voltage regulation in bus j is shown in shown Fig.2 -STATCOM location, and branches current of the network,change in the   steady-state condition. The diagram of buses *i* and *j* of the distribution systems, when D-STATCOM is installed for voltage regulation in bus. The phasor diagram of these buses with D-STATCOM placement is shown in Fig.3. The voltage of bus *j* changes from *Vj* to *Vjnew* after D-STATCOM. For simplicity, the angle of voltage *Vi* is assumed to be zero.
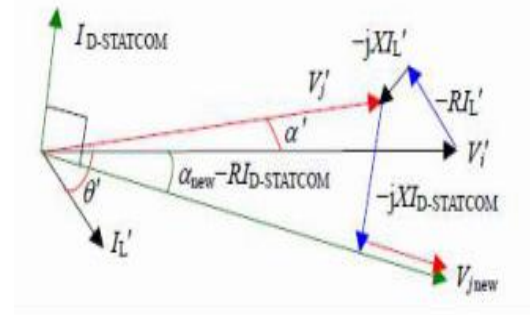
Figure 2: Single line diagram of two buses with DSTATCOM
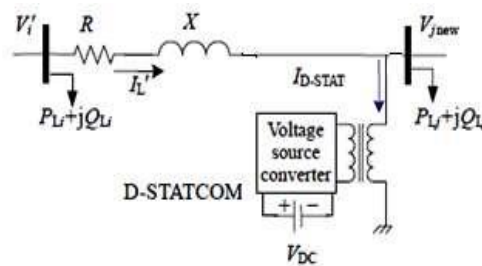
Figure 3 : Phasor diagram of voltages and currents

$V(j)\angle\alpha_{new} = Vi\angle\delta - (R+jX)IL\angle\theta - (R+jX)I_{dsat}(\angle\alpha_{new}+\pi/2)$ (15)
by equating imaginary and real parts the above equation(15) we obtain,
$Vj_{new}\cos\alpha_{new}= Re(Vi\angle\delta - Re(RI_L\angle\theta)+XI_{dsat}\sin(\angle\alpha_{new}+\pi/2)-RI_{dsat}\cos(\angle\alpha_{new}+\pi/2)$ (16)
$Vj_{new}\sin\alpha_{new}= Im(Vi\angle\delta - Im(RI_L\angle\theta)+XI_{dsat}\cos(\angle\alpha_{new}+\pi/2)-RI_{dsat}\sin(\angle\alpha_{new}+\pi/2$ (17)
$a = Re(Vi\angle\delta - Re(RI_L\angle\theta)$
$b = Im(Vi\angle\delta - Im(RX\angle\theta$
$c_1 = -R, c_2 = -X, d = Vj_{new}$
$x_1 = I_{dsat}, x_2 = \alpha_{new}$
From equations (16), (17)
$d\cos x_2 = a-c_1x_1\sin x_1-c_2x_1\cos x_2$ (18)
$d\sin x_2 = a-c_2x_1\sin x_2-c_1x_1\cos x_2$ (19)
a,b,c_1,c_2 are constants, we need to calculate x_1, x_2.
$x_1 = \dfrac{d\cos x_2-a}{-C_1\sin x_2-c_2\cos x_2}$, $x_1 = \dfrac{d\sin x_2-b}{-C_2\sin x_2-c_1\cos x_2}$ (20)
Considering $x=\sin x_2$,Equating above equation(20) we get
$(K_1^2+K_2^2)x^2+(2k_1dc_1)x+(d^2c_1^2-K_2^2) = 0$ (21)
Where,
$k_1 = a_1c_2-a_2c_1,\ k_2 = a_1c_1-a_2c_2$
The solution of equation (21) is given as,
$x = (-2k_1dc_1\pm\surd\{(2k_1dc_1)^2-4(K_1^2+K_2^2)(d^2c_1^2-K_2^2)\}$
$\alpha_{new}= x_2 = \sin^{-1}x$
Now the injected reactive power and voltage and current at node where is installed given as,
$Vj_{new} = Vj_{new}\angle\alpha_{new}$ (23)

Idsat = (∠αnew+π/2)   (24)

jQ$_{stat}$ = V$_{jnew}$I*d$_{stat}$ (25)

and * denotes conjugate of complex variable.

The DSTATCOM is modeled such that the voltage magnitude of the node where DSTATCOM is located to set to 1p.u. The phase of the node where DSTATCOM is located is calculated by using equation (22), and the current flowing in DSTATCOM i.e., Idstat is calculated from equation (24). Finally injected by DSTATCOM is calculated by equation (25).

Algorithm for radial distribution system load flow with DSTATCOM:

[1]   Read the distribution system line data and bus data.
[2]   Run the load flow of distribution system as in section 3. Select the candidate bus as in section 2.1
[3]   Assume the voltage profile of candidate bus as 1 p.u.
[4]   Calculate the injected reactive power of DSTATCOM and phase angle of the candidate bus using equations
[5]   (22), (25).
[6]   Update the reactive power and voltage phase angle of compensated bus.
[7]   Run the load flow of distribution system with updated reactive power and phase angle of candidate bus.
[8]   Stop.

IV.   LOAD MODELLING

A load model in this paper is a mathematical representation of the relationship between power and voltage, where the power is either active or reactive and the output from the model. The voltage (magnitude and/or frequency) is the input to the model. The load model could be a static or dynamic load model or a combination of both. Load models are used for analyzing power system stability problems, such as steady state stability, transient stability, long term stability and voltage control. It isnot said that the same load model is appropriate for different stability analysis. In order to obtain a model which is as simple as possible, it is important to choose a load model structure which is appropriate for the studied problem"For dynamic performance analysis, the transient and steady-state variation of the load P and Q with changes in bus voltage and frequency must be modeled" [15]. For power system analysis load can be thought as real and reactive power launched to lower voltage distribution system at buses represented as network model. Among lots of devices and appliances being connected to the system and considered as a load, one should include inter relating distribution systems feeders shunt capacitors, transformers etc. as well as voltage controlling devices or voltage regulators. The simplest load model is static model. A static load model expresses the characteristics of the load at any instant of time as algebraic functions of bus voltage magnitude and frequency at that instant. The load modelling of the active& reactive power is given by

P = P$_o$(V^a)   (26)

Q = Q$_o$(V^a)   (27)

From equations (26) & (27)

If a = 0   Constant power load model.
    a = 1   Constant current load model.
    a = 2   Constant impedance load model.

V. RESULTS AND DISCUSSION

With the help of MATLAB program, the effectiveness of the proposed VSI and DSTATCOM performance is tested on 12.66 KV, 100MVA radial distribution system consisting of 33 buses. The single linediagram of 33-bus system is shown in figure 4 and its data is given in [11].
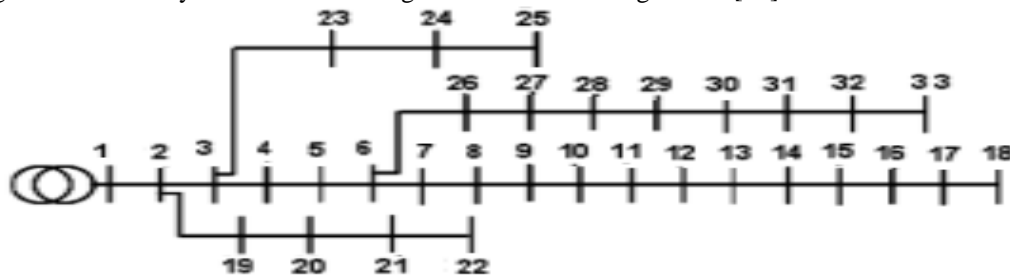
FIG.4: Single line diagram of 33-bus distribution system

The load flow of the 33 bus system is carried out and the voltages at each node and the Voltage Stability Indicator (VSI) are calculated and results shown in table below:

TABLE1: Bus voltages and VSI values from load flow

BUS	Voltage (p.u.) without DSTATCOM	Voltage Stability Indicator	Voltage (p.u.) with DSTATCOM at bus 30
1	1.0000	-	1.0000
2	0.9970	0.0003	0.9983
3	0.9830	0.0014	0.9909
4	0.9755	0.0017	0.9885
5	0.9682	0.0008	0.9863
6	0.9498	0.0015	0.9861
7	0.9463	0.0013	0.9831
8	0.9415	0.0050	0.9784
9	0.9352	0.0020	0.9726
10	0.9294	0.0020	0.9671
11	0.9286	0.0004	0.9663
12	0.9271	0.0009	0.9648
13	0.9210	0.0035	0.9591
14	0.9187	0.0028	0.9571
15	0.9173	0.0011	0.9558
16	0.9160	0.0015	0.9546
17	0.9140	0.0026	0.9528
18	0.9134	0.0024	0.9522
19	0.9965	0.0004	0.9978
20	0.9929	0.0041	0.9942
21	0.9922	0.0011	0.9935
22	0.9916	0.0019	0.9929
23	0.9794	0.0014	0.9874
24	0.9727	0.0122	0.9808
25	0.9694	0.0123	0.9775
26	0.9479	0.0004	0.9870
27	0.9453	0.0006	0.9884
28	0.9339	0.0020	1.0005
29	0.9257	0.0038	1.0002
30	0.9222	0.0298	1.0010
31	0.9180	0.0053	1.0005
32	0.9171	0.0024	1.0097
33	0.9168	0.0009	1.0095

From the TABLE 1, it is noticed that the bus 30 has the highest value of VSI. Hence Bus 30 is considered to be the weak bus. And also bus 18 has the lowest voltage. The upper and lower limit of

voltage magnitudes are 0.95 p.u. and 1.05 p.u. 21 out of 33 nodes have under voltage or over voltage problem when there is no DSATACOM installed.

In order to improve the voltage profile of the system the DSTATCOM is placed at the weak bus 30. DSTATCOM is modeled such that it injects the required reactive power to maintain the voltage at 1 p.u. at the node where it is connected and improve voltage profile of other nodes. The load flow is carried out after placingDSTATCOM at node 30. TABLE 1 shows that after DSTATCOM improves the voltage of all the nodes having under voltage problems in distribution system. The result shows that the DSTATCOM installation in this node strongly improves the voltage of neighboring nodes.

TABLE 2: Active power loss of distribution system with& without DSTATCOM

	Without DSTATCOM	With DSTATCOM at bus 30
Size of DSTATCOM (MVA)	-	3.386 MVAR
No. buses with under or over	21	0
Total Active Power losses (kw)	201.8925	143.0350
% loss reduction	-	29.15%

TABLE 2 shows the total Active power loss without DSTATCOM is 201.8925 KW, when DSTATCOM is placed at bus 30 the active power losses are reduced by 58.8575 KW. The active power loss reduction after DSTATCOM placement is 29.15%. The results in TABLE 1 and 2 are achieved based on assumption that DSTATCOM has no capacity limit for reactive power injection to voltage compensation. And the size of DSTATCOM installed at bus 30 is 3.386 MVA. In the below fig.5 shows the bus voltages in p.u. corresponding to with & without applying 3MVAR DSTATCOM at bus 30.
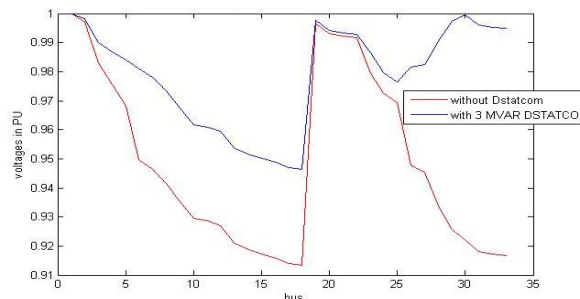

FIG.5: Voltages of with and without 3 MVAR
DSTATCOM at bus 30

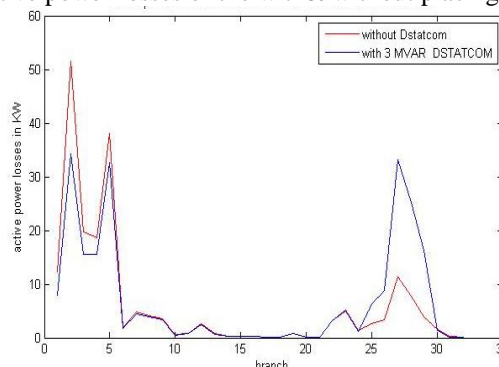The below fig.6 shows the active power losses of the with& without placing 3MVAR DSATCOM at bus 30

FIG.6: Active power losses of the with and without
3MVAR DSTATCOM at bus 30

IV. CONCLUSION

From the above simulation results, it is observed that node 30 has the highest value of VSI which is the weakest bus of 33 bus radial distribution system. Hence the DSTATCOM is placed at the weakest bus. After the placement of DSTATCOM the voltage profile of all the buses is improved. The DSTATCOM is modeled such that it maintains voltage 1 p.u. at the bus where it is installed. Hence the voltage profile of the system is improved and system losses are minimized. As an extension of the work, the load multiplier factor may be increased and check the VSI performance.The DSTATCOM in this paper is modeled for fixed voltage and no capacity limit which may result in high MVA ratings. So, our proposed idea for estimation of voltage stability indicator is good method for the operating personal to bring back the voltage level within the estimated range. The proposed estimation method is helpful in power system voltage stability limit in post contingency conditions.

REFERENCES

[1]   T.K.Abdul, G.B.Jasmon "A New Technique for Voltage Stability Analysis in a Power system & Improved Load Flow Algorithmfor Distribution Network". IEEE Catalogue no 95TH8130.

[2]   Chanda, S.; Das, B.; , "Identification of weak buses in a power network using novel voltage stability indicator in radial distribution system," International Conference on Power Electronics (IICPE), 2010 India, vol., no., pp.1-4, 28-30 Jan. 2011doi: 10.1109/IICPE.2011.5728121

[3]   Ph.D thesis of Dr. C.K. Chanda on "Global voltage stability indicator index" in 2003,BESU.

[4]   H. K. Clark, "New challenges: Voltage stability" IEEE Power Engg Rev, April 1990, pp. 33-37

[5]   T. Van Cutsem: "A method to compute reactive power margins with respect to voltage collapse", IEEE Trans. On Power Systems,No. 1, 1991

[6]   C.K. Chanda, A. Chakraborti,S.Dey, "Development of global voltage security indicator(VSI) and role of SVC on it in longitudinal

[7]   power supply(LPS) system", ELSEVIER(Electrical Power System Research 68),2004, pp.1 -9.

[8]   Bishop, M.T., Foster, J.D., Down, D.A., 1994. The Application of Single-phase Voltage Regulators on Three-phase Distribution Systems. The 38th Annual Conf. on Rural Electric Power, p.C2/1-C2/7.

[9]   Gu, Z., Rizy, D.T., 1996. Neural networks for combined control of capacitor banks and voltage regulators in distribution systems.

[10]   IEEE Trans. on Power Delivery, 11:1921- 1928. [doi:10.1109/61.544277].

[11]   Kojvic, L.A., 2006. Coordination of Distributed Generation and Step Voltage Regulator Operations for Improved Distribution System Voltage Regulation. IEEE Power Engineering Society General Meeting, p.232-237.

[12]   Ramsay, S.M., Cronin, P.E., Nelson, R.J., Bian, J., Menendez, F.E., 1996. Using Distribution Static Compensators (DSTATCOMs) to Extend the Capability of Voltage-limited Distribution Feeders. The 39th Annual Conf. on Rural Electcic power, p.18-24

[13]   Haque, M.H., 2001. Compensation of Distribution System Voltage Sag by DVR and D-STATCOM. IEEE Porto Power Tech.Conf., 1(5):223-228.

[14]   Ghosh, S., Das, D., 1999. Method for load-flow solution of radial distribution networks. IEE Proc.-Gener.Transm.Distrib.,146(6):641-648. [doi:10.1049/ip-gtd:19990464]

[15]   Ma, J., Xu, J., Wang, S., Lin, X., 2002. Calculation and Analysis for Line Losses in Distribution Network. Int. Conf. on Power System Technology, p.2537-2541.

[16]   Baran, M.E., Wu, F.F., 1989b. Network reconfiguration in distribution systems for loss reduction and load balancing. IEEE Trans.on Power Delivery, 4(2):1401-1407

[17]   M. Sedighizadeh, and A. Rezazadeh, "Load

[18]   Modeling for Power Flow and Transient Stability Computer Studies at BAKHTAR Network"

Authors:

K.NIRMALA
Pursuing PG degree in Electrical power systems at SVPCET, Affiliated to JNTUA, RVS Nagar, puttur,(AP).

 N.POORNACHANDRA RAO
Associate professor in department of EEE, SVPCET, Affiliated to JNTUA, RVS Nagar, puttur, (AP).

The Involvement of RSUs in VANETs: Survey and Perspectives

OUCHENE Faiza[1,] Boukhatem Lila[2,] GUEROUI Mourad[1]

[1] *Prism, Univ. Versailles St Quentin, Versailles France*
[2] *LRI, Univ. Paris-Sud, Orsay Cedex France*

ABSTRACT:

VANETs (Vehicular Ad hoc NETworks) are the most active domain of research in recent years. They are aimed at offering safety and non-safety applications. All type of vehicles and infrastructure are the component elements of such networks. Their important characteristics are the special mobility pattern and very dynamic topology. This leads to particular features for this type of network. Frequently, communication path is disconnected between source and destination nodes, which represent a challenge to provide routing protocols with low communication delay and low overhead. Two types of communications are present in VANET: Vehicle-To-Vehicle communication (V2V) which uses only vehicles in the process of data propagation from source to destination. And Vehicle-To-Infrastructure communication (V2I) where the infrastructures are solicited to transfer data. Several routing protocols are proposed in the literature especially in the case of V2V communication. In this paper, we mainly survey routing protocols V2I in VANETs. This study allows us to have an idea of the area of operation of RSUs in VANETs.

KEYWORDS: *Ad-Hoc Networks, Vehicular Ad Hoc Networks, Routing Protocols, V2V communications, V2I communications, Roadside units and On Board units.*

I.   INTRODUCTION

In 1998 the Intelligent Transportation Society of America adopted a definition of Intelligent Transportation Systems (ITS) [1]. It reads: People using technology in transportation to save lives, time and money. To improve safety and travel times on the transportation system, ITS exploits all the new technologies in term of electronics, telecommunications and information technology. Examples of systems concerned by ITS: traffic management, public transportation management, emergency management, traveler information, advance vehicle control and safety, commercial vehicle operations, electronic payment and railroad grade crossing safety. Several projects and activities are born to develop this technology in Europe, Japan and the USA, like carTalk [3], PreVENT [3], PRE-DRIVE [3], ASV (1, 2, 3 and 4), VSC-2, etc. VANET (Vehicular Ad hoc NETwork) consists of all types of vehicles (car, buses, etc.), referred as On Board Units (OBUs) and infrastructures installed on roads referred as Road Side units (RSUs). VANETs are dedicated to offer multiple applications; we can separate them into two main categories, safety [1] [2] and non-safety applications [3]. The first category is the main goal of VANETs, which aims to improve safety in the road by dissemination of critical alerts to vehicles. In this category, we can find accident prevention applications, collision alert, merge assistance, etc. In the second category, we find all other applications like real-time traffic congestion, high-speed tolling, mobile infotainment, online games between passengers in the road, etc. In order to implement all these applications, it is necessary to overcome the challenges related to the unique characteristics of VANETs networks. We present here the main challenges:

- **Mobility and environment conditions**
  The high speed and mobility of vehicles represent a challenge to most optimization algorithms aimed to predefine routes to forward packets.
- **Inherent characteristics of the radio channel:**
  The presence of objects in the road can degrade the quality of the received signal. In addition, the problem of the fading effects, which due to the mobility and the surrounding objects.
- **Security and privacy:**
  The users of this technology want to make sure that they can trust the source of information. Also, the privacy requirements of senders must be respected.
- **Lack of an on-line centralized management and coordination entity:**

The problem here is the efficient use of the available bandwidth of the wireless channel. In VANET, there is no entity able to synchronize and manage the transmission events of the different nodes which lead to a large number of packet collisions so a less efficient usage of the channel.
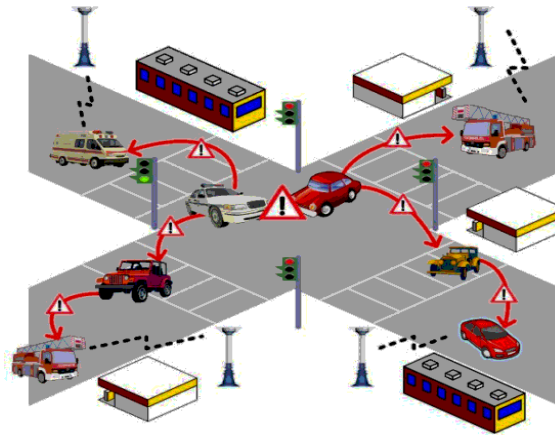
Figure 1: Different communication architectures in vehicular ad hoc networks

Figure 1 shows a structure example of vehicular networks illustrating the different communications architectures. In this figure we can distinguish two categories of communications: V2V communication represented by red continuous line, where the dotted line represents the V2I communication:

- **Vehicle-To-Vehicle communication (V2V) [1]:** The communication in this category is done through vehicles amongst themselves. Security applications are usually involved in V2V communication like warning of non-compliance with the safety distance, alert in case of accident, warning of the obstacles on the road and collaborative driving.

- **Vehicle-To-Infrastructure communication (V2I) [1] [4]:** Infrastructure is implicated in this case, in order to provide services like Internet access, information services and information on weather or traffic.

Routing protocols consist to establish routes for forwarding packets. So, for exchanging information different entities are defined. Routing protocol would be effective and ensure a good flow of packets with low delay, low overhead and good use of resources. Many routing techniques are developed to overcome the VANETs networks characteristics which are at the same time challenges. We find the following methods: *Unicast* routing which is the forwarding packet from a source to a target node in the network. *Multicast* routing delivery packet from a single source to several receivers (group of receivers called multicast group) by multi-hop communication. *Geocast* routing the packet is delivered to a specific geographic region. Vehicles of the target area should receive and forward packet. *Broadcast* routing is the process to forward packet from a source to all nodes in the network. Table 1 shows some routing protocols belonging to the previous categories.

Table 1: Some routing protocols of Vehicular ad hoc networks

Vehicular Ad hoc NETworks		
Unicast	Multicast/Geocast	Broadcast
GPCR		
VADD	DRG	DV-CAST
CAR	IVG	Broadcast Methods for V2V communications
DIR	DG-CASTOR	
Reliable routing	Mobicast routing	On the broadcast Storm Problem
GVGRID		

Different routing protocols are proposed for VANETs specifically protocols based on V2V communication. In this paper, we focus only on routing protocols V2I communication in order to identify the involvement of RSUs in the routing process in VANETs. The rest of paper is organized as follows: Section 2 is an overview of routing protocols using V2I communication. Section 3 gives some possible future perspectives and challenges for the exploitation of RSUs. And finally, we present our conclusion in the section 4.

II.   OVERVIEW OF PROTOCOLS V2I

In this section, we describe some routing protocols using the V2I communication.

2.1 Reliable Routing for Roadside to Vehicle Communications in Rural Areas

Authors in [5] propose a novel routing protocol applicable in rural roadways. This approach is based on the Access Points (APs) to maintain connections of vehicles to Internet. This protocol is proposed to address the issue of the terrain factor characteristics. This factor can be illustrated by the geographical characteristic (curve roadway and mountain) causes in the rural highway occasionally loses the line of sight between the nodes in VANET. Because of the lack of or sometimes the total absence of the fixed communication infrastructures in this area, the main solution in this case is the Multi-hop inter vehicle communication connecting to AP. Two reliable routing strategies are proposed for Roadside to Vehicle communication (R2V), prediction algorithm and routing algorithms. The first one, use the information about the current locations and velocities of nodes to predict the lifetime in terms of time units, all of this information can be piggybacked in the packets sent to APs. To establish a connection to the Internet, vehicle initiates a route discovery if it does not have a cached route. The discovery consists on flooding process by sending Route REQuest (RREQ) packets. When the request arrives to the AP, this last one selects a route for the source by sending a Route REPly (RREP). The AP can change the current path based on a tradeoff between lifetime of the path and its length. A long lifetime means a large numbers short links with long predicted lifetimes which leads to a high overhead and end up with long end-to-end. Also, the first selection of the path may have a shortest lifetime because the prediction may not be very accurate. Select a path with a minimum of hop-count leads to a quick break of links. So, the current path is changed if its predicted lifetime drops below a threshold or if the AP finds another path with the same lifetime but with smaller length. This method helps to prevent breakage of the path.

2.2 Infrastructure-Assisted Geo-Routing for Cooperative Vehicular Networks

The authors of this paper [6] propose to use the RSU for multi-hop communications to benefit of the increase of the range and reliability of communications due to the higher antenna height. The paper uses the topology-aware GSR routing protocol. Initially, with this approach, we have a road map topology where nodes are connected and the intersections are anchor points. Following the metric employed by the considered topology-aware routing protocol, the weights of the graph are calculated. They consider that the RSUs are directly connected through a network backbone. The principal of their proposition is based on the modification of the graph representation. So for the integration of the RSU in the new graph, they consider that the RSUs can be merged into a unique graph node, which is the backbone network. In the new graph "graph network", all the RSUs are represented as a unique graph node. This new representation of the infrastructure nodes will play a role in the calculation of the shortest path while vehicles perceive all RSUs as a graph node. In the final, when a RSU receives a packet, this last will be transferred to the next RSU. The new RSU will forward the packet to the node. On the other side, this technique doesn't ensure to find the short path from source to destination using a graph network that includes all the RSUs as a graph node.

2.3 A Static-Node Assisted Adaptive Routing Protocol in Vehicular Networks (SADV)

In this paper [7] a static nodes are used at road intersections to help relay data. This static nodes have a digital street map, based on which the trajectory is calculated for the packet forwarding. The path is obtained due to a graph of the street map, where we have the set of the static nodes and the set of the directed roads. The weight of each road is allocated according to the delay between adjacent static nodes. SADV is composed by three modules. SNAR (Static Node Assisted Routing): handles to store and forward data through optimal path until the disposal of vehicles to forward the packet to the next hop static node. LDU (Link Delay Update): measures the delay between intersections. To realize this, a single field is inserted in the packet head. At reception of the packet by the next static node, it will be easy to determinate the delay of the link used. This information will be encapsulate into the delay update message, a broadcast it to the others nodes by static node only. MPPD (Multi Path Data Dissemination): when the packet arrives at an intersection this last one send it to his adjacent static nodes which represent the best and the second best paths. With this technique, they increase the chance of hitting a better or even the real optimal path.

2.4An efficient routing protocol for connecting vehicular networks to the Internet

In paper [8] is destined to offer Internet connection to vehicles through a hybrid gateway discovery process to overcome the problem of high velocity of vehicles and the overhead. The communication can be established directly between vehicle and gateway if the vehicle is in the transmission range of the gateway, or in the other case through a multi-hop path. Authors assume that vehicles are equipped by a GPS in the goal to obtain their location. Speed, direction and location information will be used to predict the future location of neighbors of a vehicle. They assume also that the scenario is realized on highway. Each gateway broadcast advertisement message using Geocast capabilities in a specific area. The message contains, position, speed and direction of the sender, addresses of the relay nodes, time of the expiration of the route, zone of broadcast message and the location of the gateway, which represent the center of the circle. From the information about the distance separates gateways and the density of traffic; a zone of broadcast (circle or rectangle) is defined for each gateway. This zone delimits the process of broadcasting. When vehicles receive the broadcast message, just the vehicles located in the broadcast zone or the difference between their direction angle and the direction of the sender is less than $\pi/4$, and offers the longest lifetime, will be a relay and broadcast the message after the expiration of the timer. This mechanism reduces the problems related to the flooding of the network. The current path can be replaced with a new one if this last has the same lifetime and minimum of hops. This method insures the selection of the more stables path. Finally, a new path is found before the expiration of the lifetime of the current path due to the information collected by the advertisement messages of gateways and the method of selection of the relays.

2.5 Vertex-Based Multihop Vehicle-to Infrastructure Routing for Vehicular Ad Hoc Networks

Article [9] shows a method which permit to find a path from a vehicle to the nearest AP. Each vehicle has a digital map; this map is used by the vehicle to calculate the shortest path to the nearest AP. The path is a prediction of a sequence of intersections between source and destination (AP); this list is inserted in the packet header. Once the road found, the data transmission starts. In the road, vehicles exchange beacon messages among them, with this messages a forwarding vehicle can obtain a list of their possible future neighbors and calculate the weighted score for packet carrier of their current and future neighbors. The weight is calculated based on the position, direction and the distance between nodes and destination (Infrastructure). The neighbor with the highest weight is selected to carrier packet between intersections. In other hand, RSUs are utilized to provide short-time certificates updating for vehicles as they can be connected to the trust authority. This certificate is used to protect the privacy and security of information of users. Authors in [10] propose an algorithm for an optimal deployment of RSUs to allow the OBUs to get a new certificate before the expiration of the last one.

III.   CHALLENGES AND PERSPECTIVES

As mentioned above, VANETs are constituted of mobile nodes (OBUs) and infrastructure (RSUs). However, the exploitation of RSUs is limited to date to providing Internet access and the short time certificate. This use is restricted due to additional processing time at the RSUs, and the cost of their installation which is expensive. Therefore, some future perspectives should include the following elements:

- Introduction of RSUs in the routing process (not just for Internet access), considering them as anchors and taking advantage of their transmission coverage and their known fixed positions.
- Improvement of the dissemination of emergency messages by collecting the information through the connection to the other RSUs and mainly to the authority center.
- RSU have the potential for load balancing traffic to avoid network congestion.
- The study of optimal location methods of the RSUs in cities is very important to solve the problem of the installation cost of RSUs.

IV.   CONCLUSION

VANET is expected to play a powerful role in the improvement of safety in roads and offers more applications for the passengers during their travel in vehicle. In this paper, we presented a survey about routing protocols using V2I communications. The most works are oriented to the use of the RSU in order to provide the Internet access. RSU have the potential to provide an opportunity to improve routing in vehicular networks owing to higher antenna height which increases the range and the reliability of the V2I communications. These characteristics can be used to establish robust path for forwarding packets and they can play a role for load balancing traffic to avoid network congestion, mainly in the case of emergency messages. The use of RSU proves to be very important and deserves to be developed in order to exploit their potential in the routing process.

REFERENCES

[1] http://www.its.dot.gov/
[2] http://www.itscanada.ca/it/index.html.
[3] http://www.vanet.info/
[4] http://www.vehicle-infrastructure.org/
[5] S. Wan, J. Tang, and R.S. Wolff. Reliable Routing for Roadside to Vehicle Communications in Rural Areas. Proceedings of International Conference on Communications, pp. 3017 - 3021, Beijing, May 19-23, 2008.
[6] D. Borsetti and J. Gozalvez, "Infrastructure-Assisted Geo-Routing for Cooperative Vehicular Networks", IEEE Vehicular Networking Conference , pp. 255 - 262 Jersey City, NJ, December 13-15, 2010.
[7] Y. Ding, C. Wang and L. Xiao. A Static-Node Assisted Adaptive Routing Protocol in Vehicular Networks. IEEE Transactions on Vehicular Technology - IEEE TRANS VEH TECHNOL (Volume: 59, Issue: 5), 2010, pp. 2445 - 2455.
[8] A. Benslimane, S. Barghi and C. Assi. An efficient routing protocol for connecting vehicular networks to the Internet. Pervasive and Mobile Computing, pp. 98-113,February, 2011.
[9] R. K. Shrestha, S. Moh, I. Chung and D. Choi. Vertex-Based Multihop Vehicle-to-Infrastructure Routing for Vehicular Ad Hoc Networks. 43rd Hawaii International Conference on Systems Science, Koloa, Kauai, HI, USA, pp. 1-7, 5-8 January 2010..
[10] Y .Sun, R. Lu, X. Lin, J. Su and X. Shen. Wolff. Roadside Units Deployment for Efficient Short-time Certificate Updating in VANETs. Proceedings of International Conference on Communications, pp. 1-5, Cape Town, May 23-27, 2010.

DESIGN AND ANALYSIS OF DS-CDMA DETECTED MULTIPATH SIGNALS USING THE RAKE RECEIVER SIMULATOR FOR WIRELESS COMMUNICATION.

[1,] Parisae.Veera Swamy, [2,] M Hari Krishnam Raju [3,]P.Suresh

[1,2,3,]*M.Tech student,Asst. Prof.,ECE HOD*
Dept. of ECE,Sri Sunflower College of Engg.&Technology
Lankapalli(Challapalli), Krishna (Dt), A.P - 521131

ABSTRACT
Wireless cellular telephony has been growing at a faster rate than wired-line telephone networks. This growth has also been fueled by the recent improvements in the capacity of wireless links due to the use of multiple access techniques (which allow many users to share the same channel for transmission) in association with advanced signal processing algorithms. Code Division Multiple Access (CDMA) is becoming a popular technology for cellular communications. One form of CDMA called Direct Sequence CDMA (DS-CDMA) uses a set of unique signature sequence or spreading codes to modulate the data bits of deferent users.

I.   INTRODUCTION

The number of multipath signals in the wireless channel is unknown and difficult to predict. The spread spectrum technology aims to spread the information signal over a wider bandwidth to make jamming and interception more difficult. , the number of multipath signals in the wireless channel is unknown and difficult to predict. A RAKE receiver allows each arriving multipath signal to be individually demodulated and then combined to produce a stronger and more accurate signal. The RAKE receiver in the IS-95A CDMA system uses three correlators and a searcher, while the TIA/EIA-95B CDMA system limits the number of correlators in the RAKE receiver to six. The searcher receives pilot signals for synchronizing the spreading code. Both of these systems have a fix number of correlators and leads to the RAKE receivers either containing an excessive number of correlators or that the receiver performs sub optimally.

II.   CDMA

Code Division Multiple Access (CDMA) is a spread spectrum technique that uses neither frequency channels nor time slots. In CDMA, the narrow band message (typically digitized voice data) is multiplied by a large bandwidth signal, which is a pseudo random noise code (PN code). All users in a CDMA system use the same frequency band and transmit simultaneously. The transmitted signal is recovered by correlating the received signal with the PN code used by the transmitter.
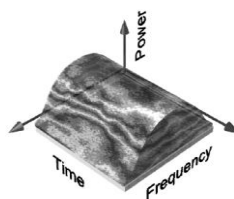
Fig: Code Division Multiple Access (CDMA)
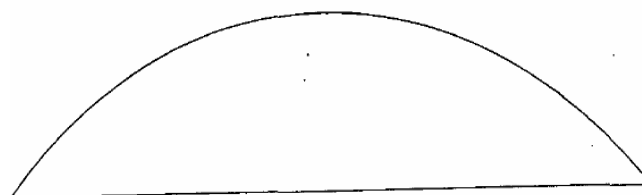
Fig: Spectrum of Wideband CDMA

2.1.Direct-Sequence Spread Spectrum(DS-SS)

The DS-SS technique is one of the most popular forms of spread spectrum. This is probably due to the simplicity with which direct sequencing can be implemented. Fig 4.13 shows the basic model and the key characteristics that make up the DS-SS communications system. In this form of modulation, a pseudo-random noise generator creates a spreading code or better known as the pseudo-noise (PN) code sequence. Each bit of the original input data is directly modulated with this PN sequence and is represented by multiple bits in the transmitted signal. On the receiving end, only the same PN sequence is capable of demodulating the spread spectrum signal to successfully recover the input data.
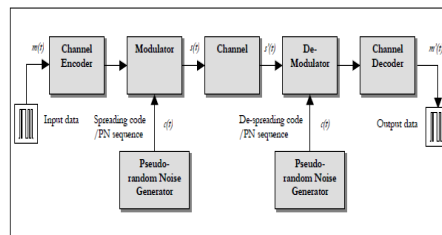
Fig: Basic Model of the Direct-Sequence Spread Spectrum Communications System

The bandwidth of the transmitted signal is directly proportional to the number of bits used for the PN sequence. A 7-bit code sequence spreads the signal across a wider frequency band that is seven times greater than a 1-bit code sequence, otherwise termed as having a processing gain of seven. Fig 4.14 illustrates the generation of a DS-SS signal using an exclusive-OR (XOR) operation. The XOR obeys the following rules:

$$0 \oplus 0 = 0 \qquad 0 \oplus 1 = 1 \qquad 1 \oplus 0 = 1 \qquad 1 \oplus 1 = 0$$
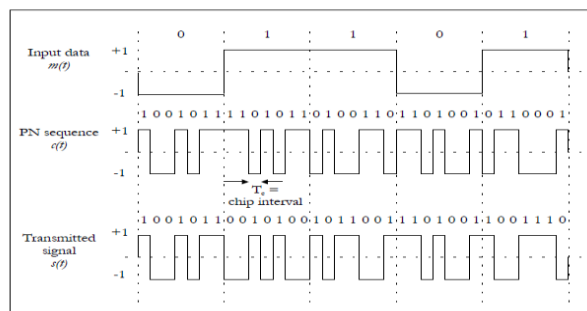
Fig: Generation of a DS-SS Signal with Processing Gain = 7

III.   RAKE RECEIVER

If in a mobile radio channel reflected waves arrive with small relative time delays, self interference occurs. Direct Sequence (DS) Spread Spectrum is often claimed to have particular properties that makes it less vulnerable to multipath reception. In particular, the rake receiver architecture allows an optimal combining of energy received over paths with different. It avoids wave cancellation (fades).if delayed paths arrive with phase differences and appropriately weighs signals coming in with different signal-to-noise ratios.The rake receiver consists of multiple correlators, in which the receive signal is multiplied by time-shifted versions of a locally generated code sequence. The intention is to separate signals such that each finger only sees signals coming in over a single (resolvable) path.

Fig: Simple Block Diagram of DS-CDMA Transmitter and Receiver

IV. SIMULATION

4.1.GUI for simulation of CDMA

When we run the rake_cdma_gui.m then this will open this gui the simulation will perform on this gui this GUI show the information about the no of bit error in simulation with and without rake , show the user data and received data, user have to put information about the no of user , user to simulate and attenuation factor and after simulation we will get figure of transmit data and received data , number of bit error in received data for both RAKE and without RAKE receiver , this result show the comparison between simulation with and without RAKE receiver in cdma
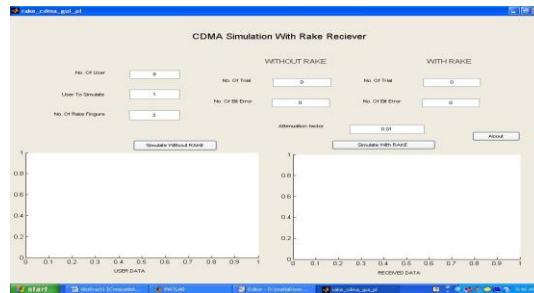
Fig: GUI for Simulation of CDMA

4.2.GUI for CDMA simulation without RAKE when attenuation factor is 1

When we push the push button simulation without RAKE button with attenuation factor 1 then the transmitting data and receiving data, no of bit error , the attenuation factor is one in this condition bit error is zero this show that the power level of transmitting data is high then the bit error in receiving data is less.
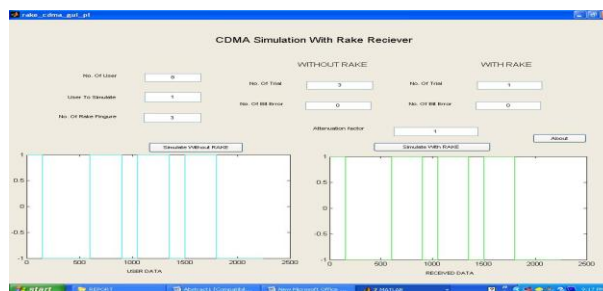
Fig: GUI for CDMA Simulation without RAKE When Attenuation Factor is 1

4.3.GUI for CDMA simulation with RAKE when attenuation factor is 1

When we push the push button simulation with RAKE button with attenuation factor is 1 then the transmitting data and receiving data, no of error no of trial . the attenuation factor is one in this condition bit error is zero this show that the power level of transmitting data is high then the bit error in receiving data is less.
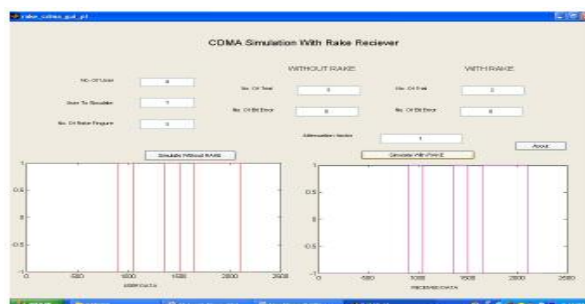
Fig: GUI for CDMA Simulation with RAKE Reciever

4.4.Multipath received signal

When we push the push button simulation without RAKE button or simulation with RAKE with attenuation factor is 1 then the multipath receiving data.
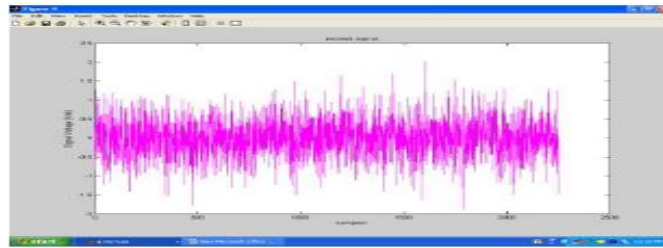
Fig: Multipath Received Signal

4.5Command window after simulation with RAKE

Command window of Matlab after simulation with RAKE will show information about the receiving data err1 is the receiving data which we get from path one and err2 is the receiving data which we get from path two.

Fig: Command Window after Simulation with RAKE

4.6.Command window after simulation with RAKE

Command window of Matlab after simulation with RAKE will show information about the receiving data the err3 is the receiving data which we get from path three the mxco1, mxco2 and mxco3 is the total sum of err1, err2 and err3 we will take maximum output of three.
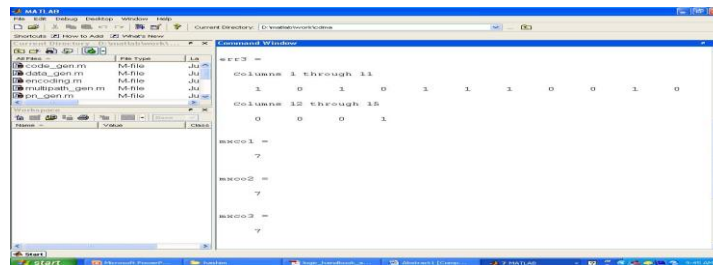
Fig: Command Window after Simulation with RAKE

4.7.Command window after simulation without RAKE

Command window of Matlab after simulation without RAKE will show information about the receiving data the deco_data is the output data and err is show the bit error in receiving data.
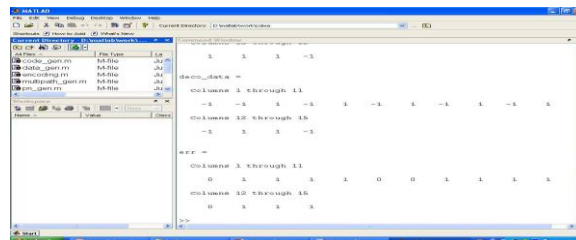
V.   CONCLUSION

We have developed a simulator made in our project, we have shown how rake receiver in used for CDMA to decease bit error due to multipath interferences. it can simulate a CDMA encoding and decoding process, the data is assumed to be travelled through different path, the effect of different path and CDMA is

considered to generate the multipath effect the data is pass through the different path and assumed at receiver end just before the decoder .to avoid the multipath effect RAKE receiver is used the RAKE receiver concept is introduced in the decoding process.Detection of DS-CDMA signals using the RAKE Rx, results in performance improvement compared to correlator Rx.

VI. REFERENCE

[1] Peter Flanagan, "Personal Communications Services: The Long Road Ahead," Telecommunications, February 1996.
[2] http://wireless.per.nl/reference/chaptr05/cdma/rake.htm.
[3] W. C. Y. Lee, "Overview of Cellular CDMA,"IEEE Trans. On Vehicular Technology, Vol. 40, no. 2, pp. 291-302, May 1991.
[4] R. A. Cameron and B. D. Werner, "An Analysis of CDMA with Imperfect Power Control," Proceedings of 42nd IEEE Vehicular Technology Conference, Denver, CO, pp. 977-980, 1992.
[5] R. Lupus and S. Verdi, "Linear Multiuser Detectors for Synchronous Code Division-Multiple-Access Channels," IEEE Trans. Info. Theory, vol. 35, no.1, pp. 123-136, Jan. 1989.
[6] K.Murali.Krishna, Abhijit Mitra and Cemal Ardil" A Simplified Single Correlator Rake, Receiver for DMA Communications" International Journal of Information Technology Volume 2 Number 4 2005.
[7] P. Jung, P. W. Baier, and A. Steil, "Advantages of CDMA and Spread Spectrum Over FDMA and TDMA in Cellular Mobile Radio Applications," IEEE Transactions Vehicular Technology,Vol. 42, no. 3, pp. 357- 364, August 1993
[8] Electronic Industries Association, "Cellular System Dual-Mode Mobile Station Base Station Compatibility Standard," IS-54, May 1990.
[9] K. S. Gilhousen, "On the Capacity of a Cellular CDMA System,"IEEE Transactions on Vehicular Technology, Vol. 40, no. 2, pp. 303-311,May 1991.
[10] Electronic Industries Association, "Widband Spread Spectrum Digital Cellular System Dual Mode Mobile Station - Base Station Compatibility Standard," IS-95, April, 1992
[11] J. C. Liberti and T. S. Rappaport, "Analytical Results for Capacity Improvements in CDMA," IEEE Transactions on Vehicular Technology, Vol.43, No. 3, pp. 680-690, August 1994.

A Triband Slotted Bow-Tie Antenna for Wireless Applications

Dr. Siva Agora Sakthivel Murugan[1], K.Karthikayan[2], Natraj.N.A[3], Rathish.C.R[4]

Associate Professor[1], Assistant Professor[2,3,4]
[12]Department of EEE, [3]Department of ECE,
Sree Sakthi Engineering College[1,2], United Institute of Technology[3,4]

ABSTRACT

This paper presents a new approach for the design of a tri-band bowtie antenna. By incorporating slots with triangular shapes on the arms of the bowtie, resonance is obtained in the 2.5 GHz, 4.4 GHz, and 6.2 GHz bands. A study of the effect of triangular slots, their size and their location on the return loss of the antenna is presented. The designed antenna enjoys advantages such as low profile, low cost of fabrication, and high radiation efficiency. Details of the design along with experimental and simulation results in FEKO are presented and discussed. .The simulation results of proposed antenna are analyzed by using Method of Moment (MoM) from FEKO software.

INDEX TERMS: *Triband,Triangular slots,Resonance.*

I. INTRODUCTION

A simple bowtie antenna is made from a bi-triangular sheet of metal with the feed at its vertex. This type of antenna is used extensively in many applications such as ground penetrating radar and mobile station.Bowtie antennas have many advantages such as low profile, high radiation efficiency, ease of manufacturing and low fabrication cost. With many IEEE certified wireless networking standards are booming it is expected several bandwidth will be made available for open use. The techniques used to obtain multi-resonance were borrowed from patch antennas . In [8], two parallel rectangular slots are incorporated into the antenna patch, leading to 30% increase in bandwidth. The [2]resulting E-shaped patch antenna had another desirable property: it had two resonant frequencies. In this work, triangular slots are introduced on each arm of the bowtie antenna, leading to triple-resonance which, by modifying the location and size of the slots.Due to the rapid progress in wireless communication systems, high gain broadband antennas are of great demand. Bow-tie slot antennas have the advantage of wide bandwidth.

Antennas performance and size have a large impact on the development of wireless system. Compared to traditional antenna it is more complicated to provide the typical parameters like bandwidth, efficiency and gain within the limited antenna volume. This becomes even more critical with respect to the UWB system with high data rate and low power density. Microstrip antennas are extensively used in different applications due to their many advantages, such as low profile, light weight, easy to fabricate and low cost and also higher bandwidth. Because of the limited bandwidth of the patch antennas,several designs of broadband patch antenna have been studied, some of them are [1]-[6]. The radiation characteristics of the printed dipole antenna expected to be similar with rectangular patch except for those features that depend on the width to length ratio. The input impedance, bandwidth and the cross polarization radiation can be different extensively. In this paper a new broad-band planar impedance matching scheme is achieved by using a simple triangular slots without any structural complexities. The shape and position of the triangular slot is changed . So more electromagnetic energy is coupled in to the patch .The antenna has three working frequencies and can be used in wireless applications.

II. DIFFERENT BOWTIE DESIGNS

In this section, bowtie antennas designed with different slot sizes are presented. By studying the effect eachconfiguration has on the reflection coefficient of the antenna allows us to decide on the final design that produces resonance in the 2.5GHz, 4.4GHz and 6.2GHz bands
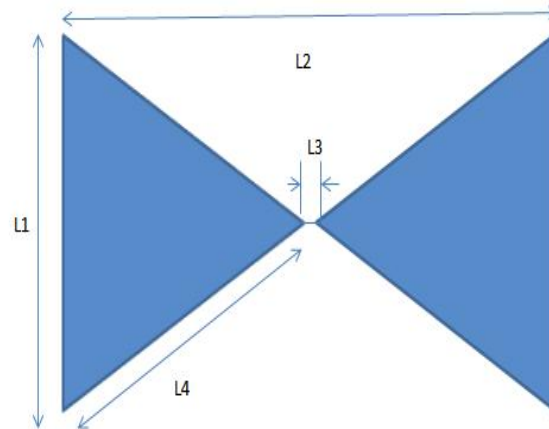
Fig 1:A simple Bow-Tie antenna

A bowtie antenna is shown in Fig. 1. In this case, the antenna L1 is 17.2 mm ,L2 is 32.2,L3 is .5mm and L4 15.2. The corresponding antenna is working in 2.4Ghz frequency . In order to obtain multiband antenna the geometry of this antenna is modified trying various slots and patches of various shapes.For a triangular slot triband antenna a triangle of similar dimension is placed inside the border of existing antenna The inner and outer triangles making each slot are isosceles and share the same vertex. The dimension of the antenna is L1 23.88 mm,L2 34.42mm,L3 .5mm,L4 33.02mm,L5 34.44 and L6 .11mm. The dimensions of the arms are the same used in the first design. The S11 plot of this antenna is shown in Fig. 3. Therefore, the incorporation of triangular slots will give resonance in the bands of interest.The dimensions of the arms of the bowtie were changed in order to keep the impedance matched. On each, triangular slots of equal dimensions were taken. Both triangles share the same vertex, with an opening width of 0.5 mm.A line fed is given at the meeting point which drives the antenna.
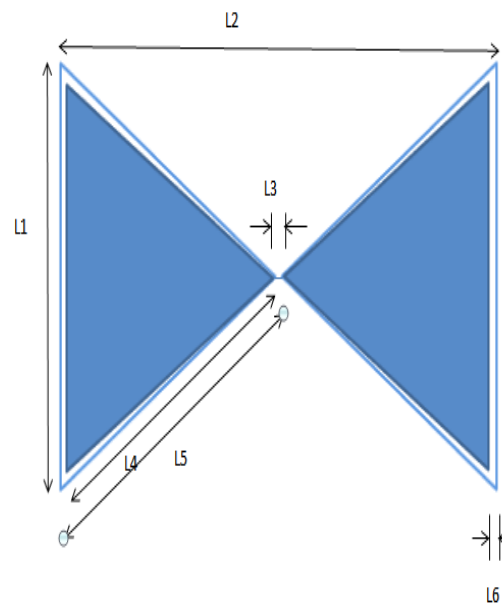
Figure 2: Geometry of the proposed antenna

III.   RESULTS AND DISCUSSIONS

The simulation and the experimental studies of the antenna are done using FEKO suit 6 which is basically a recent tool used for electromagnetic analysis involving bodies of arbitrary shape. Fig. 3 shows the simulated and experimental return loss characteristics of the antenna. Return loss is achieved minimum at three major resonances centered at 2.5 GHz, 4.4 GHz and 6.2 GHz respectively .
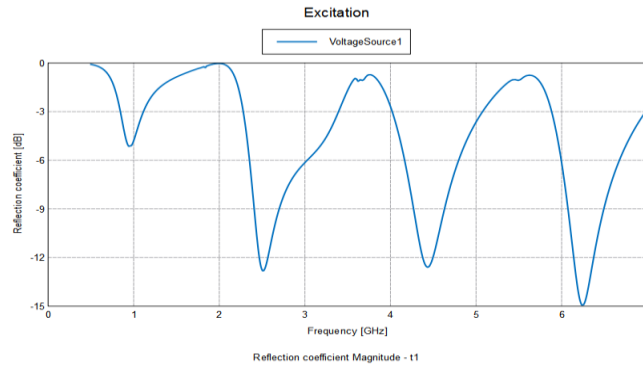
Figure 3: Return loss characteristics of proposed antenna

The patches and slots are placed on several places on bowtie antenna and resultant graphs are analyzed and superior performance is observed in the proposed model.The spacing between the triangular patch is varied in order to find a better impedance matched circuit.

Fig. 4 shows the return loss characteristics of the antenna with different patch dimensions . It is evident from the graph that with triangular patch of higher dimension there exists only two poorly matched resonances. As length of the spacing increases impedance matching also increases and the maximum bandwidth is obtained when L6=.11mm. Also the resonant frequencies shift towards the lower side with increase in L6.
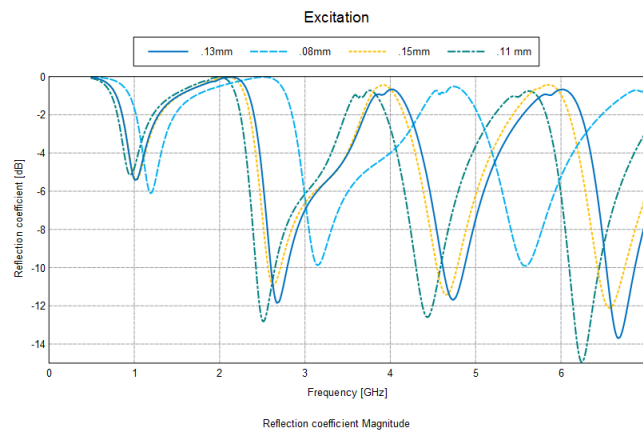
Fig 4: plot of reflection coefficient characteristics of proposed antenna with varying L6

It is clear from the below figure the simulated voltage standing wave ratio (VSWR) of the modified bow-tie antenna is less than 2 for the fixed frequencies 2.5 GHz,4.4 GHand6.2GHz   for other frequencies(VSWR) is greater than two. The resulting return loss plot for this antenna is shown in Fig.3. Triple resonance is achieved: at 2.5GHz (– 12.8dB), at 4.4 GHz (- 12.6dB) and at 6.2 GHz (-14.7 dB). The antenna radiation pattern at Φ= 00 is shown in Fig. 5.
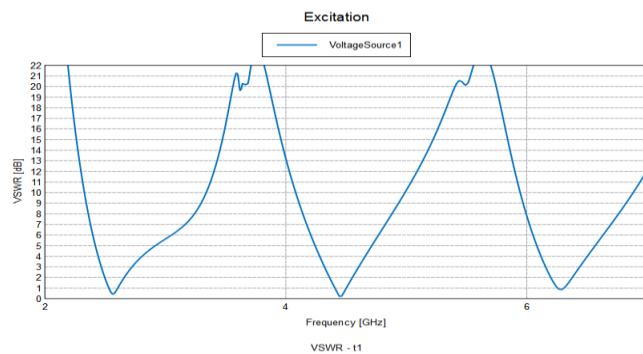
Figure 5:VSWR characteristics of proposed antenna
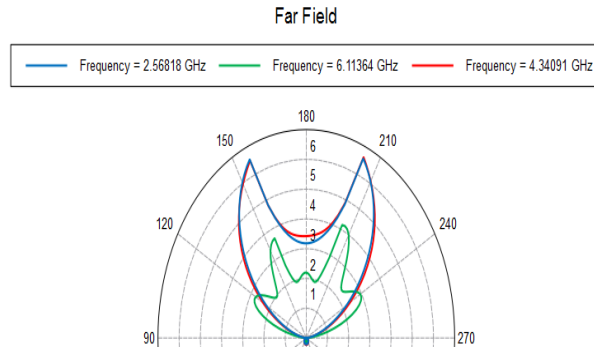
(a)radiation pattern with elevation(theta=0,phi=90)
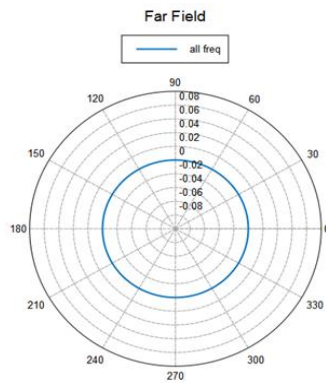
(b)radiation pattern with azimuthal
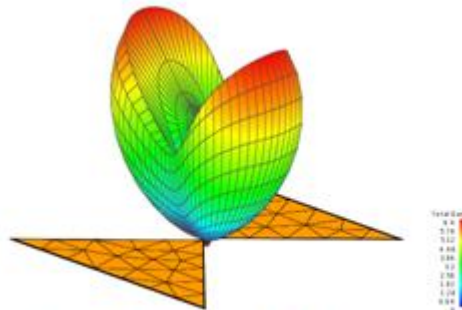
(c) 3D far field at 2.4 GHz

(d)Far field pattem at 4.4GHz
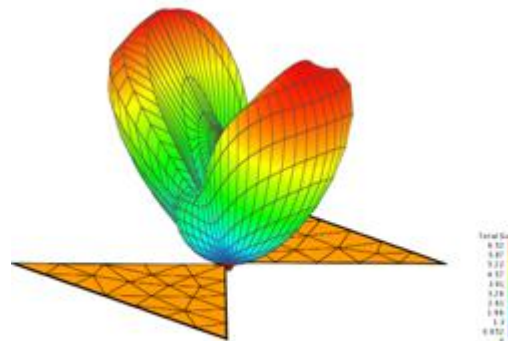
(e)Far field pattern at 6.2GHz
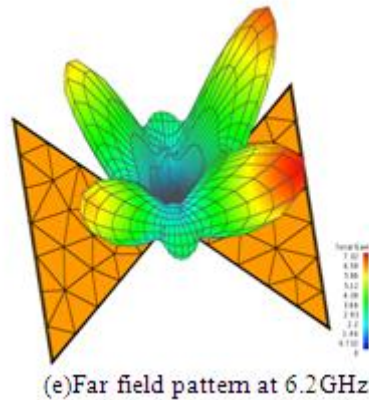
Figure 6: simulated far field patterns in2D and 3D

IV.   CONCLUSION

The above figures shows the simulated radiation patterns with Elevation and azimuthal at different frequencies by using CAD FEKO software. The simulated radiation patterns of antenna in the E-plane (XZ-plane) and H-plane (YZ-plane) for three different frequencies 2.5 GHz, 4.4GHz and 6.2 GHz are shown in figure.6 . The patterns and other curves are obtained at the time of simulation. We observed good radiation patterns by taking 20 cells per wavelength.

REFERENCES

[1]   Uduwawala, D. et al., "A deep parametric study of resistor-loaded bowtie antennas for ground penetrating radar applications using FDTD,"IEEE Trans. Geosciences and Remote Sensing, vol.. 42, No. 4, pp. 732-742, Jun. 2004.
[2]   Y. Nishioka, O. Maeshima, T. Uno and S. Adachi. ,"FDTD analysis of resistor-loaded bowtie antennas covered with ferrite-coated conducting cavity for subsurface radar," IEEE Trans. Antennas and Propagation, vol. 47, No. 6, pp. 970-977, Jun. 1999.
[3]   Birch, M. and Palmer K. D., "Optimized bowtie antenna for pulsed lowfrequency ground-penetrating radar," Proceeding of SPIE, vol. 4758,2002.
[4]   Y. Lin. and S. Tsai, "Analysis and design of broadside-coupled striplines-fed bowtie antennas," IEEE Trans. Antennas and Propagation,vol. 46, No. 3, pp. 459-460, Mar. 1998
[5]   Olexa R., Implementing 802.11, 802.16, and 802.20 Wireless Networks.Oxford: Elsevier Inc., 2005.
[6]   Cherry S. M., "WiMAX and Wi-Fi: Separate and Unequal," IEEE Spectrum, March 2004.
[7]   C.A. Balanis, "Antenna Theory, Analysis and Design", John Wiley &Sons, 2005.
[8]   Yang F. Et al., "Wide-band E-shaped patch antennas for wireless communications," IEEE Trans. Antennas and Propagation., vol. 49, No.7, July 2001

An Enhanced Localization Scheme for Mobile Sensor Networks

Dr. Siva Agora Sakthivel Murugan[1], K.Karthikayan[2], Natraj.N.A[3], Rathish.C.R[4]

Associate Professor[1], Assistant Professor[2,3,4]
[12]DepartmentofEEE, [3]DepartmentofECE,
Sree Sakthi Engineering College[1,2], United Institute of Technology[3,4]

ABSTRACT

Localization in mobile sensor networks is more challenging than in static sensor networks because mobility increases the uncertainty of nodes positions. The localization algorithms used in the Mobile sensor networks (MSN)are mainly based on Sequential Monte Carlo (SMC) method. The existing SMC based localization algorithms commonly rely on increasing beacon density in order to improve localization accuracy and suffers from low sampling efficiency and also sampling in those algorithms are static and have high energy consumption. Those algorithms cannot able to localize sensor nodes in some circumstances.The main reason for that is in some time slots the sensor node cannot hear any beaconnode. This results in localization failure. The Improved Monte Carlo Localization (IMCL) algorithm achieves high sampling efficiency, high localization accuracy even in the case when there is a low beacon density. This can be achieved using bounding box and weight computation techniques. This algorithm also uses time series forecasting and dynamic sampling method for solving the problem of localization failure. Simulation results showed that the proposed method has a better performance in sparse networks in comparison with previous existing method.

KEYWORDS : *Mobile Sensor Networks, Localization,Sequential Monte Carlo methods.*

I.   INTRODUCTION

Wireless sensor networks (WSNs) have been used in many fields, including environmental and habitat monitoring, precision agriculture, animal tracking, and disaster rescue. In many applications, it is essential for nodes to know their positions. For example, data should be labelled with the positions where they are collected to help the scientists perform corresponding analysis. Position information of nodes are also necessary in many network protocols, e.g., clustering and routing which depend on the geographical information of nodes. The procedure through which the nodes obtain their positions is called localization. In localization, the nodes in a sensor network can be categorized into two types: beacon nodes which are aware of their positions and sensor nodes which need to determine their positions using a localization algorithm.

A straightforward method for localization in WSNs is to use existing localization techniques, e.g., attaching a Global Positioning System (GPS) receiver on every sensor node. However, as the scale of sensor networks becomes larger and larger, these methods become infeasible because of their high cost or inconvenience. In some recently emerging applications such as animal monitoring and tracking sensor nodes may move after deployment. These nodes form mobile sensor networks in contrast to traditional static sensor networks in which sensor nodes remain stationary after deployment. The motion of sensor nodes makes most existing localization algorithms designed for static sensor networks inapplicable to mobile sensor networks. There are some localization algorithms specially designed for mobile sensor networks, All of them are based on the Sequential Monte Carlo (SMC) method. This is because the SMC method provides simple simulation-based approaches in estimating the location. Previous SMC-based localization algorithms either suffer from low sampling efficiency or require high beacon density to achieve high localization accuracy. The major problem of most existing SMC-based localization algorithms is that they only rely on increasing beacon density to improve localization accuracy. However, beacon nodes are usually more expensive than sensor nodes. Because there are much more sensor nodes than beacon nodes in a sensor network, it will be very beneficial if sensor nodes can be used to improve the localization accuracy.In this paper, we propose an efficient algorithm which addresses both aforementioned issues. The algorithm is based on the sequential Monte Carlo Localization (MCL) algorithm named as Improved MCL (WMCL). IMCL achieves high sampling efficiency and achieves high localization accuracy even when the beacon density is low using bounding box technique and weight computation

methods.Despite the above technique haves good localization accuracy, sampling in these techniquesare static and they have high energy consumption. Also the existingalgorithms are not able to localize sensor nodes in somecircumstances. The main reason is that in some time slots the node cannot hear any seed node. TheImproved Monte Carlo Localization (IMCL) algorithm uses forecasting and dynamic sampling method forlocalization. This method has the ability of nodes localization in those conditions and it is an energy efficient method.The paper is organized as follows. The Section 2 deals with related work in localization of Wireless Sensor Networks. Section 3 deals with the proposed Improved Monte Carlo Localization (IMCL) algorithm. Section 4 is devoted to extensive performance analysis. Section 5 deals with the Conclusion and future directions.

II. RELATED WORK

Extensive research has been done on localization for wireless networks. A general survey is done focusing only on localization techniques suitable for ad hoc sensor networks. The approaches taken to achieve localization in sensor networks differ in their assumptions about the network deployment and the hardware's capabilities. Centralized localization techniques depend on sensor nodes transmitting data to a central location, where computation is performed to determine the location of each node. Doherty, Pister and Ghaoui developed a centralized technique using convex optimization to estimate positions based only on connectivity constraints given some nodes with known positions. MDS-MAP technique improves on these results by using a multidimensional scaling approach, but still requires centralized computation. Requiring central computation would be infeasible for mobile applications because of the high communication costs and inherent delay, hence we focus on distributed localization techniques.Distributed localization methods do not require centralized computation, and rely on each node determining its location with only limited communication with nearby nodes. These methods can be classified as range-based and range-free. Range-based techniques use distance estimates or angle estimates in location calculations, while a range-free solution depends only on the contents of received messages. Range-based approaches have exploited time of arrival, received signal strength, time difference of arrival of two different signals (TDOA), and angle of arrival (AOA). Though they can reach fine resolution, either the required hardware is expensive (ultrasound device for TDOA, antenna arrays for AOA) or the results depend on other unrealistic assumptions about signal propagation (for example, the actual received signal strengths of radio signals can vary when the surrounding environment changes). Because of the hardware limitations of sensor devices, range-free localization algorithms are a cost effective alternative to more expensive range-based approaches.Monte Carlo localization (MCL) method is developed for use in robotics localization for use in mobile sensor network applications. MCL is a particle filter combined with probabilistic models of robot perception and motion. It outperforms other proposed localization algorithms in both accuracy and computational efficiency. The key idea of MCL is to represent the posterior distribution of possible locations using a set of weighted samples. Each step is divided into a prediction phase and an update phase. In the prediction phase, the robot makes a movement and the uncertainty of its position increases. In the update phase, new measurements (such as observations of new landmarks) are incorporated to filter and update data. The process repeats and the robot continually updates its predicted location.

However, there are substantial differences between robot localization and node localization for sensor networks. While robot localization locates a robot in a predefined map, localization in sensor networks works in a free space or unmapped terrain. Second, a robot has relatively good control and probabilistic knowledge of its movement in a predefined map. A sensor node typically has little or no control of its mobility, and is unaware of its speed and direction. Third, a robot can obtain precise ranging information from landmarks, but a sensor node can only learn that it is within radio range. Finally, in robot localization, the individual measurements are integrated multiplicatively, assuming conditional independence between them, and the weights of samples need to be normalized after updating. In MCL, due to the constraints in computing and memory power, a filtering approach is adopted in which each measurement can be considered independently, and the weight of each sample is either 0 or 1. There are some other variants of MCL, for example, the dual and Mixture MCL, Multihop-based Monte CarloLocalization (MMCL), and Range-based MCL. The dual and Mixture MCL improves the localization accuracy of MCL by exchanging the probability functions used in the sampling step and in the filtering step. It incurs higher computational cost than MCL. MMCL and Range-based MCL use multihop sensor-beacon distances to improve the localization accuracy and to reduce the number of needed beacons. Compared with them, Our WMCL algorithm doesn't use multihop sensor-beacon distances so incurs much less communication cost and it uses weight computation methods to minimize localization error.

The previous existing methods have two major weak points which are not focused. The first problem is using of constant number of samples for localization. Second is that all nodes in all time slots cannot be localized. The proposed methods will overcome those drawbacks and the methods are discussed below.

III. THE PROPOSED IMCL ALGORITHM

A. Introduction

In IMCL algorithm, the network model is introduced and then the five main parts of IMCL are described as follows:Bounding-box construction, Dynamic sampling, Time series forecasting, Samples weights computing, maximum possible localization error computing.

B. Building the Bounding Box

There are two areas involved in bounding box of IMCL: the candidate samples area and the valid samples area. The candidate samples area is used to draw new candidate samples and the valid samples area is used to filter out invalid samples. When the candidate samples area is large and the valid samples area is small, candidate samples drawn in the sampling step have high probability to be filtered out in the filtering step.figure1 shows the construction of bounding box in IMCL algorithm.
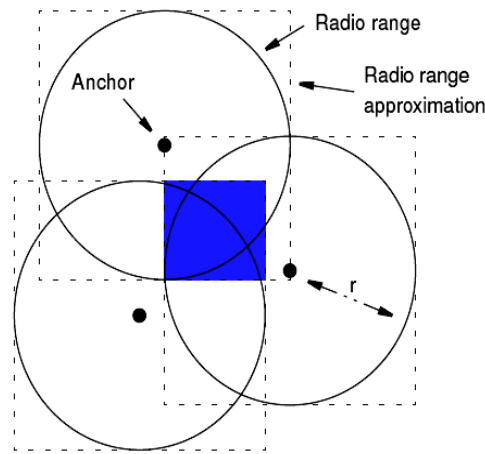
Figure1: Building the Bounding-box

In IMCL, the possible locations of a sensor node after move lie in a disk with radiusV_{max}. So the size of the candidate samples area will increase when V_{max} increases. On the other hand, when sd increases, the size of the valid samples areas will decrease. Denote by V_t the total number of candidate samples drawn in the sampling step in time unit t and define the sampling efficiency in t as,

$$e_t = \frac{|L_t|}{vt} \quad (1)$$

then in WMCL the sampling efficiency will decrease when V_{max} or Sd increases, which will cause high computational cost accordingly. Two-Hop beacon neighbours are also used to reduce the size of the bounding-box by replacing r with 2r. The candidate samples are chosen from the bounding box.

B. Dynamic sampling method

As previously mentioned, instead of taking a fixed number of samples like 50 for localization, we can determine this number dynamically based on the size of the sampling area. It is clear that for a large anchor box, a large number of samples are needed to estimate nodes location accurately. While in the case of small anchor box, we will focus on a small area. For a small area, small number of samples is needed to accurately estimate nodes position.

If created anchor box have the coordinates of (X_{min}, Y_{min}) and (X_{max}, Y_{max}), so area size of this box is determined and we will specify the number of samples based on this area size. Noting to the standard number of samples that is 50, this number will be used for an anchor box with maximum area size. An anchor box is maximized when the node hears only a one-hop anchor node. In this case the box size will be equal to a square of size $2V_{max}$. So we will consider 50 samples for this box and use equation (5) for other box sizes.

$$\text{Sample Number}=50*((x_{max}-x_{min})(y_{max}-y_{min}))/_{4V_{max}}{}^2 \quad (2)$$

When the anchor box has the maximum area, numerator and denominator of the fraction in the equation (2) will be equal and so the number of samples will be equal to 50.

For the cases when the anchor box size is more than $4V_{max}^2$, we consider the number of samples equal to 50. Example of such cases is when the sensor node hears only one two-hop anchor node.

C. **Linear prediction using time series :** Linear prediction method is a powerful technique for predicting time series in a time-varying environment. Thismethod is expressed in equation(6) and is a recursive method

$$y(t + T) = a_1y(t)+a_1y(t-T)+ ... +a_my(t-(m-I)T) \qquad (3)$$

Estimated value at time t as a linear function of previous values in the times "t-T, t-2T. .. t-mT" has been produced is obtained. In equation(3) a1, a2, ..., am are the linear prediction coefficients, 'm' is the model degree, 'T' is the sampling time, y(t+T) is the next observation estimation and y(t), y(t-T), ..., y(t-mT) are the present and past observations. The prediction error which is the difference between the predicted and the real locations (Equation (4))must be minimized.

$$Error(\%)= \{\frac{predictedlocation - Reallocation}{Reallocation}\}* 100\% \qquad (4)$$

In order to estimate the coefficients of linear prediction model we use the least squares error method and rewrite equation (3) with considering modelling error in equation (4):

$$y(t) = a_1y(t)+a_1y(t-T)+ ... +a_my(t-(m-I)T) +e (t) \qquad (5)$$

The error e(t) is generated because of not adopting the linear prediction model to the real value. So to find the coefficients, a1, a2, …… am in equation(5), we use the sum least squares error and set of linear functions presented in equation(6)

$$\begin{bmatrix} y(t) \\ y(t-T) \\ M \\ y(t-mT) \end{bmatrix} = \begin{bmatrix} y(t) & y(t-2T) & \Lambda & y(t-mT) \\ y(t-2T) & & \Lambda & y(t-(m+1)T) \\ M & & M & \\ y(t-(k+1)T) & & \Lambda & y(t-(m+k)T) \end{bmatrix} * \begin{bmatrix} a_1 \\ a_2 \\ M \\ a_m \end{bmatrix} + \begin{bmatrix} e(t) \\ e(t-T) \\ M \\ e(t-kT) \end{bmatrix} \qquad (6)$$

$$Y=\phi+A+E \qquad (7)$$

Elements in the matrix A are the coefficients which can be found by least squares error method in equation (8):

$$A=(\phi^T \phi)^{-1} \phi^T Y \qquad (8)$$

In equation (8), ϕ^T is the transpose of the matrix ϕ and $(\phi^T \phi)^{-1}$ is the inverse of matrix. After obtaining the coefficients a1, a2 ... am, the nodes location in the next time slot predicted using equation (3). If the node do not localized using WMCL algorithm with dynamic sampling, we will use this predicted location instead. Then the weights are computed for predicted samples.

D. **Weighting the Samples :** After a sample candidate is chosen, its weight is computed using 1-hop and 2-hop (anchor and common) neighbour nodes. Figure2 shows the phenomenon of weight computation.
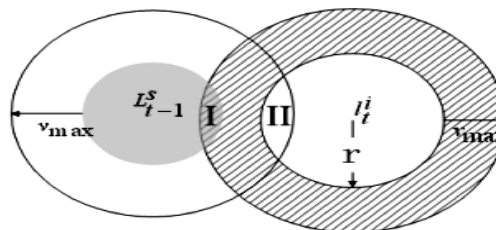
Figure 2: Weight Computation Method

In WMCL, $O_{t =} S \cup T \cup US$. So the weight of a candidate sample is computed as

$$w_t= p(O_t|l_t) = \prod_{s \in S,T} p(s | l_t) \prod_{s \in US} p(s | l_t) \qquad (12)$$

When $s \in S$ or $s \in T$, P(s $| l_t$) can be easily computed
If $s \in S$, then

$$P(s \mid l_t) = [d(l_t,s) \le r] \qquad (13)$$

If $s \in T$, then

$$P(s \mid l_t) = [r < d(l_t,s) \le 2r] \qquad (14)$$

The weights of samples are computed. The samples with zero weights are rejected and samples with high weights are taken for error computation.

E. Error Computation

After obtaining N valid samples, a sensor node computes the weighted average of these samples as its position estimation. Using the position estimation and the bounding- box, a sensor node can compute its ER_x and ER_y, as illustrated in Figure3. A more riskily method is to use the smallest rectangle enclosing all of p's valid samples to compute ER_x and ER_y. This method can improve localization accuracy a lot. However, when using this method the procedure of constructing the bounding-box should be carefully manipulated. In this case the inequality p causes some inconsistence in the computation.
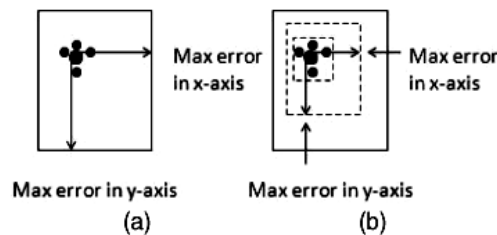
Figure 3: Computing maximum localization error

For example, it is possible that x_{min} is larger than x_{max} and consequently the bounding-box cannot be built. After p gets (x_e,y_e) and ER_x, ER_y, it broadcasts them to its neighbors. Its neighbors will use this information to compute their position estimation in the next time unit.Algorithm of the proposed method has been presented in figure (4).

Figure 4: IMCL Algorithm

IV. RESULTS AND DISCUSSION

The simulation is carried out in NS-2 simulator under Linux platform with simulation area of 1000 x 1000m and 150 mobile nodes.The node is randomly placed from there onwards the node mobility occurs in random direction. The simulation time is 100 seconds.

A. Sampling efficiency
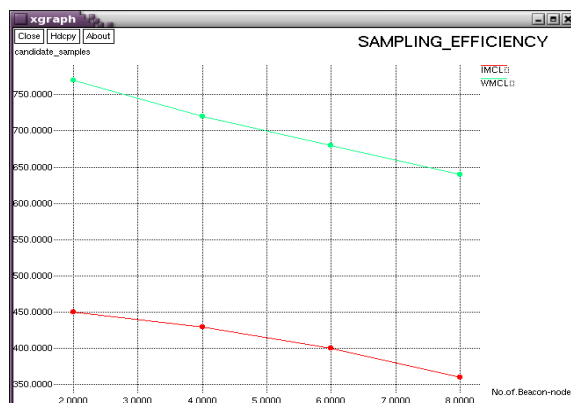
Figure 5: Samples Vs Beacon nodes
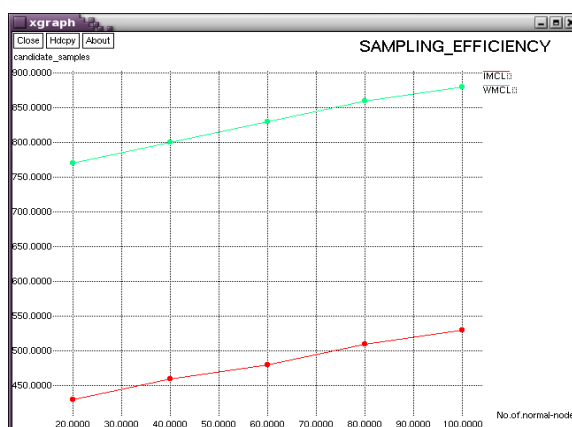
Figure 6: Samples Vs Sensor nodes

The sampling efficiency is a very import metric in SMC-based localization algorithms because higher sampling means less candidate samples generation and consequently less computational cost. Figure (5) & (6) shows the number of candidate samples obtained forlocalization in the WMCL and IMCL algorithms with varying number of beacon and sensor nodes. Thenumber of samples in WMCL algorithm is fixed andeach unknown node during each time slot uses 50 samples to do localization. But in the IMCL algorithm samples number isdetermined dynamically. Simulation results show that with using IMCL algorithm, less candidate samples are chosen thereby obtaining high sampling efficiency.each unknown node uses fewer samples than other methods.

B. Sampling Attempts
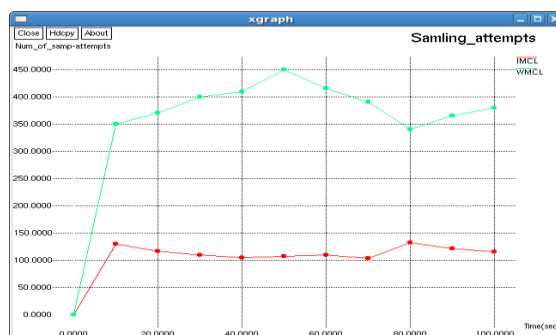
Figure 6: Sampling attempts

Most of computational energy consumption for these algorithms is related to the number of used samples and the number of sampling attempts to produce acceptable samples. Also the response time depends on the number of sampling attempts for production of required samples. Figure (6) showssimulation results for the number of sampling attempts to produce enough valid samples.

C. Dynamic Sampling
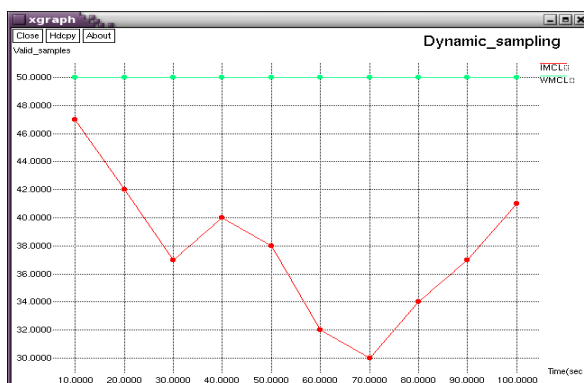
Figure 7: Dynamic sampling

The Dynamic sampling is performed in mobile nodes. Figure (7) shows the number of valid samples obtained for localization in the WMCL and IMCL algorithms. The number of samples in WMCL algorithm is fixed and each unknown node during each time slot uses 50 samples to do localization. But in the IMCL algorithm samples number is determined dynamically. Simulation results show that with using IMCL algorithm, each unknown node uses fewer samples than WMCL method.

D. Localization Accuracy

Localization accuracy is the most important metric in evaluating localization algorithms. The localization accuracy is determined from the estimated value of localization error. The localization error is noted for different time periods. The localization error is also determined by varying number of beacon and sensor nodes.
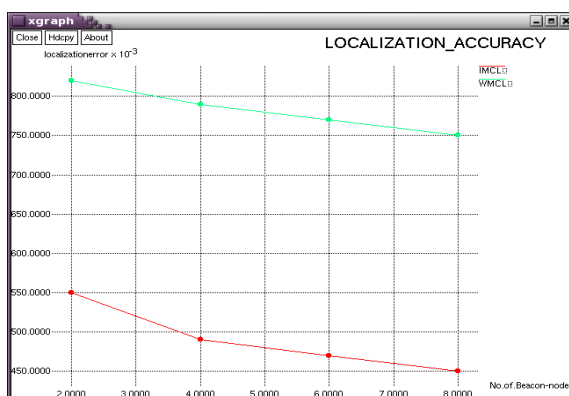
Figure 8: Localization errorVs Beacon nodes
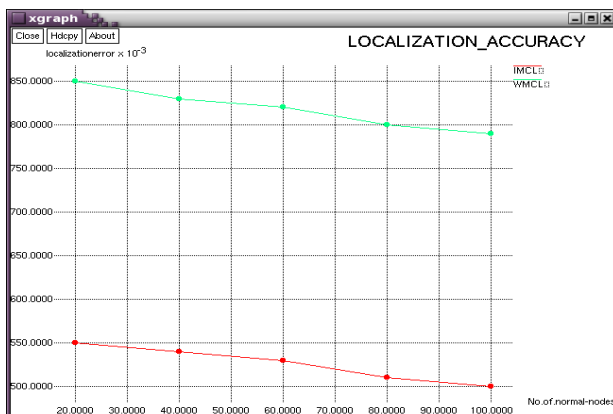
Figure 9: Localization error Vs sensor nodes

Figure(7) & (8) shows the localization error for varying number of beacon and sensor nodes. The graph shows that the localization error in IMCL algorithm is reduced compared with WMCL algorithm and thereby having high localization accuracy.

E. Energy backlog
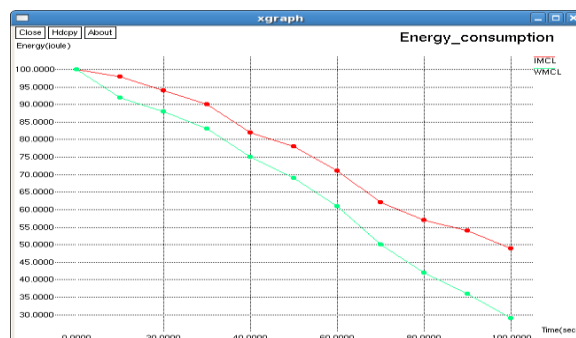
Figure 10: Energy backlog

The dynamic sampling method and TSF method used in IMCL algorithm will reduce the energy consumed in localization process by minimizing the number of sampling operation. Fig 10shows the amount of energy consumed in IMCL algorithm is very much reduced than the WMCL algorithm.

V. CONCLUSION

The Improved Monte Carlo Localization (IMCL) algorithm achieved high sampling efficiency, high localization accuracy even in the case when there is a low beacon density using the bounding box and weight computation methods. The localization accuracy is improved by using the estimated position information of sensor nodes. The proposed IMCL algorithm used dynamic sampling based on the size of sampling area to estimate the sensor nodes position. Also the proposed algorithm uses TSF method to predict sensors position when the sensor nodes do not hear any anchor nodes. The proposed algorithm is suitable for mobile sensor networks with low anchor node density. This algorithm has less implementation costs in comparison with previous method.Simulation results showed that the proposed algorithm providesbetter performance than the similar method in the sparse sensor networks.

REFERENCES

[1] Baggio and Langendoen, K "Monte-Carlo Localization for Mobile Wireless Sensor Networks", Proc. Conf. Mobile Ad-Hoc and Sensor Networks (MSN '06), pp. 317-328.2006.
[2] B. Soltaninasab and J.Amiri, "Improving Monte Carlo Localization Algorithm Using Time Series Forecasting Method and Dynamic Sampling in Mobile WSNs," Proc. Conf. Communication Systems, Networks and Applications, pp.389-396, 2010.
[3] Boukerche .A, H. Oliveira, E. Nakamura , and A. Loureiro, "Secure localization algorithms for wireless sensor networks," *IEEE Commun. Mag.,* vol. 46, no. 4, pp. 96-101.2008.
[4] Datta, S. Klinowski, C. Rudafshani, M. and Khaleque, M "Distributed Localization in Static and Mobile Sensor Networks,"Proc. IEEE Int'l Conf. Wireless and Mobile Computing, pp. 69-76.2006.
[5] Hu, L and Evans, D "Localization for Mobile Sensor Networks,"Proc. ACM MobiCom, pp. 45-57.2004.
[6] Tilak. S, V. Kolar, N. B. Abu-Ghazaleh, and K.-D. Kang, "Dynamic localization protocols for mobile sensor networks ," in *Proceedings of MASS, 2004.*
[7] T.-H. Lin, P. Huang, H.-H. Chu, and C.-W. You, "Energy-efficient boundary detection for rf-based localization systems," *IEEE Trans. Mobile Comput.,* vol. 8, no. I, pp. 29-40, 2009.
[8] Hill .J, R. Szewczyk, A. Woo, S. Hollar, D. E. Culler, and K. S. J. Pister,"System architecture directions for networked sensors," in *Proceedings ofASPLOS,* 2000, pp. 93- 104.
[9] Karp, B and Kung, T "GPSR: Greedy Perimeter Stateless Routing for Wireless Networks," Proc. ACM MobiCom, pp. 243-254.2000.
[10] Lim, H and Hou, J "Localization for Anisotropic Sensor Networks "Proc. IEEE INFOCOM, pp. 138-149.2005.
[11] Shigeng, Z. Cao, J. Lijun, C and Daoxu, C "Accurate and Energy-Efficient Range-Free Localization for Mobile Sensor Networks", IEEE transactions on mobile computing, Vol. 9.pp.897-909.2010
[12] Shigeng, Z. Cao, J. Lijun, C and Daoxu, C"Location Nodes in Mobile Sensor Networks More Accurately and Faster," Proc. Ann.IEEE Conf. Sensor and Ad Hoc Comm. and Networks.2008.
[13] Wang, W and Zhu, Q "Varying the Sample Number for Monte Carlo Localization in Mobile Sensor Networks," Proc. IEEE Int'l Multi-Symp. Computer and Computational Sciences, pp. 490-495.2007.
[14] Zeng. Y, J. Cao, S. Zhang, S. Guo, and L. Xie, "Pollution attack: A new attack against localization in wireless sensor networks ," in *Proceedings of WCNC.2009.*
[15] Zeng. Y, S. Zhang, S. Guo, and X. Li, "secure hop-count based localizationin wireless sensor networks," in *Proceedings of CIS.2007.*
[16] Zeng. Y and J. Cao, "SecMCL: A Secure Monte Carlo Localization Algorithm for Mobile Sensor Networks ," Proc . Conf. Mobile Ad-Hoc and Sensor Networks,2009

Analysis&Optimization of Design Parameters of Mechanisms Using Ga

B.Venu[1], Dr.M.nagaphani sastry[2]

[1]Student, M.Tech (CAD/CAM), G.pullareddyengineering college(Autonomous), A.P, India, [2]Associate professor, mechanical engineering, G.pullareddyengineering college(Autonomous), A.P, India,

ABSTRACT

The main objective of this study is to investigate of dynamic reaction forces of a crank mechanism. Therefore, this study consists of three major sections: (1) dynamic reactions investigation, (2) analysis of the mechanisms (3) optimization for static analysis. Analysis on slider crank mechanism is performed to calculate the reaction forces. This data is implemented for regression analysis for regression equation. These parameters are aimed to be optimized using GA. Because genetic algorithm is give good optimal values comparing to traditional optimization. This traditional optimization was done by using MATLAB.

KEYWORDS: dynamic reactions, regression analysis, genetic algorithm (GA), MATLAB.

I.   INTRODUCTION

Here mechanism is a slider-crank mechanism. The slider-crank mechanism is one of the most useful mechanisms in modern technology since it appears in most of the internal combustion engines including automobiles, trucks and small engines. The slider-crank kinematic chain consists of four bodies linked with three cylindrical joints and one sliding or prismatic joint. It is used to change circular into reciprocating motion, or reciprocating into circular motion.
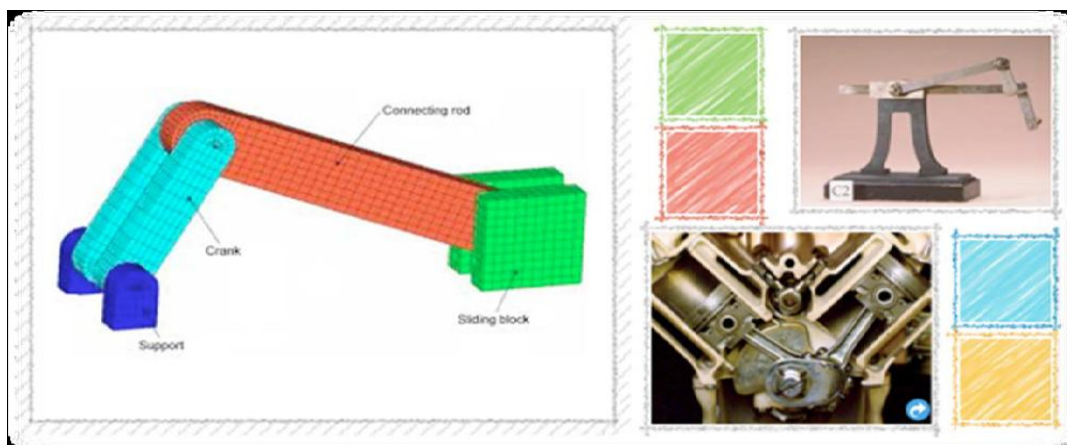
Figure 1: Slider Crank

The arm may be a bent portion of the shaft, or a separate arm attached to it. Attached to the end of Velocity analysis of slider crank mechanism the crank by a pivot is a rod, usually called a connecting rod. The end of the rod attached to the crank moves in a circular motion, while the other end is usually constrained to move in a linear sliding motion, in and out.

A mechanism is used to produce mechanical transformations in a machine. This transformation could be any of the following.
- It may convert one speed to another speed.
- It may convert one force to another force.

- It may convert one torque to another torque.
- It may convert force into torque.
- It may convert one angular motion to another angular motion.
- It may convert angular motion into linear motion.
- It may convert linear motion into angular motion.

1.1 STUDY OBJECTIVES :-

- Determine all loads acting on the links in a mechanism to allow stress and deflection analysis.
- Determine input torque(s) required to produce desired motion in a mechanism(input torque = torque supplied by input device)
- This present study in the design of machine elements includes the minimization of weight of the individual components in order to reduce the over all weight of the machine elements.
- It saves both cost and energy involved.
- The most important problem that confronts practical engineers is the mechanical design, a field of creativity.
- Mechanical design can be defined as the selection of materials and geometry, which satisfies the specified and implied functional requirements while remaining within the confines of inherently unavoidable limitations.

1.2MAT LAB :-
Here we can calculate the dynamic reactions of a slider crank mechanism by using MATLAB. MATLAB is an abbreviation for MATrix LABoratory. It is a matrix-based system for scientific calculations. we can solve numerical problems without necessarily having to write a long pro-gram. This course provides an introduction to MATLAB. It will provide the basics of MATLAB programming and applications (primarily) for macroeconomics and international finance. MATLAB is a high-level language and interactive environment for numerical computation, visualization, and programming. Using MATLAB, we can analyze data, develop algorithms, and create models and applications. The language, tools, and built-in math functions are enable to explore multiple approaches and reach a solution faster than with spreadsheets or traditional programming languages, such as C/C++ or Java. we can use MATLAB for a range of applications, including signal processing and communications, image and video processing, control systems, test and measurement, computational finance, and computational biology. More than a million engineers and scientists in industry and academia use MATLAB, the language of technical computing.

1..2.1Genetic algorithm :
The Genetic Algorithm and Direct Search Toolbox is a collection of functions that extend the capabilities of the Optimization Toolbox and the MATLAB® numeric computing environment. The Genetic Algorithm and Direct Search Toolbox includes routines for solving optimization problems using
•Genetic algorithm
•Direct search
These algorithms are enabling to solve a variety of optimization problems that lie outside the scope of the standard Optimization Toolbox. All the toolbox functions are MATLAB M-files, made up of MATLAB statements that implement specialized optimization algorithms. we can view the MATLAB code for these functions using the statement

type function _ name
we can extend the capabilities of the Genetic Algorithm and Direct Search Toolbox by writing our own M-files, or by using the toolbox in combination with other toolboxes, or with MATLAB or Simulink®.

Dynamic reaction forces on MATLAB :
```
%1lb=453.592grams
g=386.4;
wbd=5.5*453.592; %weight of the connecting rod
wp=6.3*453.592; %weight of the piston
mp=wp/g; %mass of the piston
mbd=wbd/g;
l=10; %length of the connecting rod
b=3.5; %crank radius
i_bar=(1/12)*mbd*l^2; %mass moment of inertia
```

```
omega_AB=1000*(2*pi)/60;
v_B=b*omega_AB;
theta=[0:10:180];
t=theta*pi/180;
beta=asin(b*sin(t)/l);
omega_BD=v_B*cos(t)./(l*cos(beta));
%acceleration
a_B=b*omega_AB^2;
alpha_BD=(l.*omega_BD.^2.*sin(beta)-a_B.*sin(t))./(l.*cos(beta));
a_D=a_B.*cos(t)+l.*omega_BD.^2.*cos(beta)+l.*alpha_BD.*sin(beta);
%
ax_bar=-0.5*a_B*sin(t);
ay_bar=0.5*a_B*cos(t)+0.5*a_D;
Dy=-mp*a_D;
Dx=-Dy.*tan(beta)+(i_bar*alpha_BD)./(l*cos(beta))-mbd*ax_bar./2+mbd*ay_bar.*tan(beta)./2;
Bx=mbd*ax_bar+Dx;
By=mbd*ay_bar-Dy;
%
%determine and plot values
z=[theta;Bx;By;Dx;Dy;ax_bar;ay_bar];
fprintf('theta Bx    By    Dx    Dy    ax_bar   ay_bar\n')
fprintf('(deg) (gr)  (gr)  (gr)  (gr)  (m/s^2)  (m/s^2)\n')
fprintf('\n')
fprintf('%5.0f %5.0f %5.0f %5.0f  %5.0f  %5.0f   %5.0f\n',z);
fprintf('\n');
figure(1)
plot(theta,Bx,theta,By)
xlabel('theta(degrees)')
ylabel('dynamic reactions(gr)')
legend('Bx','By')
grid on
%
figure(2)
plot(theta,Dx,theta,Dy)
xlabel('theta(degrees)')
ylabel('dynamic reactions(gr)')
legend('Dx','Dy',2)
grid on
%
figure(3)
subplot(2,1,1);
plot(theta,ax_bar);
xlabel('theta(degrees)')
ylabel('x_acceleration(m/s^2)')
grid on
subplot(2,1,2);
plot(theta,ay_bar);
xlabel('theta(degrees)')
ylabel('y_acceleration(m/s^2)')
grid on
```

Out put:

Theta (deg)	Bx (gr)	By (gr)	Dx (gr)	Dy (gr)	ax_bar (m/s^2)	ay_bar (m/s^2)
0	0	674377	0	-383202	0	0 45099

10	17488	658456	39003	-373429	-3332	44146
20	30081	611562	72458	-344676	-6564	41337
30	33591	536343	95543	-298668	-9595	36812
40	25255	437293	104899	-238340	-12336	30815
50	4419	320762	99335	-167850	-14701	23684
60	-27010	194739	80294	-92437	-16620	15845
70	-64614	68217	51818	-17987	-18034	7780
80	-102289	-49923	19733	49731	-18899	-30
90	-133761	-152351	-9856	106057	-19191	-7170
100	-154287	-234567	-32265	148312	-18899	-13360
110	-161754	-295461	-45322	176180	-18034	-18475
120	-156707	-336923	-49402	191416	-16620	-22537
130	-141470	-362729	-46554	197065	-14701	-25659
140	-118990	-377260	-39346	196549	-12336	-27989
150	-91942	-384522	-29989	192980	-9595	-29967
160	-62319	-387635	-19942	188794	-6564	-30798
170	-31414	-388714	-9899	185653	-3332	-31451
180	0	-388947	0	184505	0	-31665

Plots:
theta(deg) vs dynamic reaction(gr)
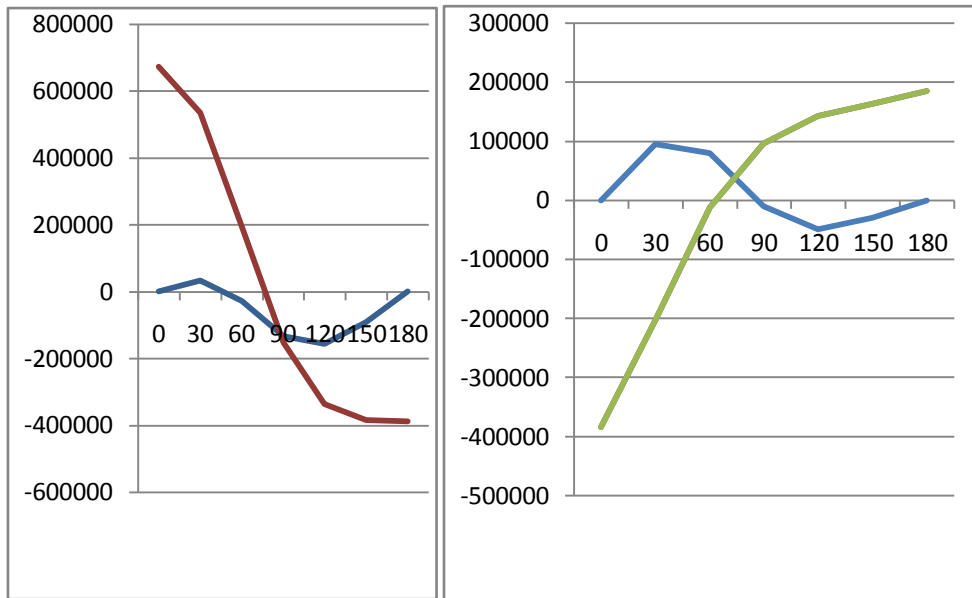
2.Response surface optimization of slider crank mechanism:-

Std	run	Factor 1 A:A	Response 1 R1	Response 2 R2	Response 3 R3	Response 4 R4
1	1	0	0	674377	0	-383202
2	2	10	17488	658456	39003	-373429
3	3	20	30081	611562	72458	-344676
4	4	30	33591	536343	95543	-298668
5	5	40	25255	437293	104899	-238340
6	6	50	4419	320762	99335	-167850
7	7	60	-27010	194739	80294	-92437
8	8	70	-64614	68217	51818	-17987
9	9	80	-102289	-49923	19733	49731
10	10	90	-133761	-152351	-9856	106057
11	11	100	-154287	-234567	-32265	148312
12	12	110	-161754	-295461	-45322	176180
13	13	120	-156707	-336923	-49402	191416

14	14	130	-141470	-362729	-46554	197065
15	15	140	-118990	-377260	-39346	196549
16	16	150	-91942	-384522	-29989	192980
17	17	160	-62319	-387635	-19942	188794
18	18	170	-31414	-388714	-9899	185653
19	19	180	0	-388947	0	184505

2.1 ANOVA table:

Response 1:

Source	Sum of Squares	df	Mean square	F value	p-value prob>F	
Model	9.037E+010	6	1.506E+010	7759.79	<0.0001	significant
A-A	1.888E+010	1	1.888E+010	9725.85	<0.0001	
A^2	4.921E+009	1	4.921E+009	2535.55	<0.0001	
A^3	3.638E+009	1	3.638E+009	1874.39	<0.0001	
A^4	1.022E+009	1	1.022E+009	526.47	<0.0001	
A^5	7.862E+008	1	7.862E+008	405.04	<0.0001	
A^6 RESIDUAL	4.082E+008 2.329E+007	1 12	4.082E+008 1.941E+006	210.32	<0.0001	
COR TOTAL	9.039E+010	18				

Obser vations:

I. The model F-value of 7759.79 implies the model is significant. There is only a 0.01% chance that a "model F-value" this large could due to noise.

II. Values of " prob>F " less than 0.0500 indicate model terms are significant.

III. In this case A,A^2,A^3,A^4,A^5,A^6 are significant model terms.

IV. Values greater than 0.1000 indicate the model terms are not significant.

V. If there are many insignificant model terms(not counting those required to support hierarchy), model reduction may improve the model.

R-Squared Results:

Std.Dev	1393.17	R-squared	0.9997
Mean	-59774.89	Adj R-squared	0.9996
C.V.%	2.33	Pred R-squared	0.9968

PRESS	2.903E+008	Adeq precision	232.752

The " pred R-squared"of 0.9968 is in reasonable agreement with the " adj R-squared" of 0.9996.
"adeq precision" measures the signal to noise ratio. A ratio greater than 4 is desirable. The ratio of 232.752 indicates an adequate signal. This model can be used to navigate the design space.
Model equation of response1:-

R1=   +121.51151+748.60025*A+103.03328*A^2-4.26162*A^3+0.048831*A^4-2.25466E-004*A^5+3.74055E-007*A^6.

Proceed to diagnostics plots(the next icon in progression). Be sure to look at the:
1. Normal probability plot to the studentized residuals to check for normality of residuals.
2. Studentized residuals versus predicted values to check for constant error.
3. Externally studentized residuals to look for outliers, i.e., influential values.
4. Box-Cox plot for power transmissions
   If all the model statistics and diagnostic plots are OK, finish up with the model graphs icon.
   R1 vs A:A
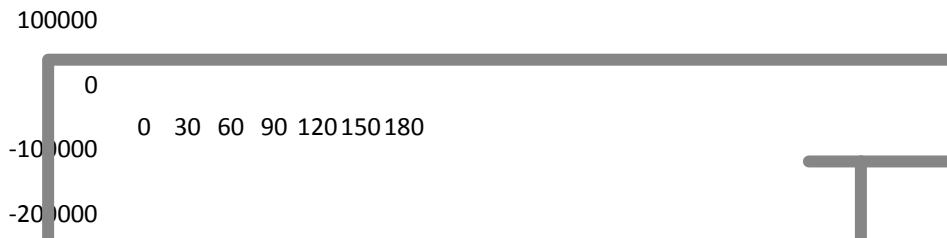
Response 2:

Source	Sum of Squares	df	Mean square	F value	p-value prob>F	
Model	3.042E+012	6	5.069E+011	5.474E+005	<0.0001	significant
A-A	2.433E+011	1	2.433E+011	2.628E+011	<0.0001	
A^2	1.870E+010	1	1.870E+010	20199.67	<0.0001	
A^3	3.088E+009	1	3.088E+009	3334.99	<0.0001	
A^4	2.414E+009	1	2.414E+009	2606.43	<0.0001	
A^5	5.986E+007	1	5.986E+007	64.64	<0.0001	
A^6	5.380E+008	1	5.380E+008	581.06	<0.0001	
RESIDUAL	1.111E+007	12	9.260E+005			
COR TOTAL	3.042E+012	18				

Obser vations:
1. The model F-value of 547448.75 implies the model is significant. There is only a 0.01% chance that a model F-value" this large could due to noise.
2. values of " prob>F " less than 0.0500 indicate model terms are significant.
3 In this case A,A^2,A^3,A^4,A^5,A^6 are significant model terms.
4. Values greater than 0.1000 indicate the model terms are not significant.
5 If there are many insignificant model terms(not counting those required to support hierarchy), model reduction may improve the model.

R-Squared Results:

Std.Dev	962.28	R-squared	1.0000
Mean	7511.42	Adj R-squared	1.0000
C.V.%	12.81	Pred R-squared	1.000
PRESS	1.037E+008	Adeq precision	1822.836

The " pred R-squared"of 1.0000 is in reasonable agreement with the " adj R-squared" of 1.0000 "adeq precision" measures the signal to noise ratio. A ratio greater than 4 is desirable. The ratio of1822.836 indicates an adequate signal. This model can be used to navigate the design space.

Model equation of respone2:-

R2= +6.74815E+005-613.87632*A-95.42986*A^2-2.34427*A^3+0.044055*A^4-2.38374E-004*A^5+4.29437E-007*A^6.

Proceed to diagnostics plots(the next icon in progression). Be sure to look at the:
1. Normal probability plot to the studentized residuals to check for normality of residuals.
2. Studentized residuals versus predicted values to check for constant error.
3. Externally studentized residuals to look for outliers, i.e., influential values.
4. Box-Cox plot for power transmissions
   If all the model statistics and diagnostic plots are OK, finish up with the model graphs icon.
   R2 vs A:A
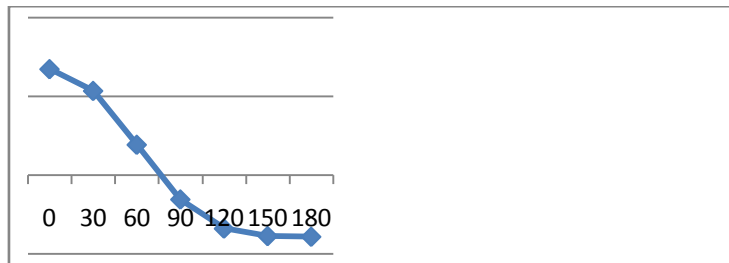
Response 3:

Source	Sum of Squares	df	Mean Square	F value	p-value prob>F	
Model	5.287E+010	6	8.812E+009	4544.83	<0.0001	significant
A-A	1.888E+010	1	1.888E+010	9736.05	<0.0001	

A^2	1.910E+009	1	1.910E+009	985.14	<0.0001	
A^3	3.638E+009	1	3.638E+009	1876.36	<0.0001	
A^4	8.906E+008	1	8.906E+008	459.35	<0.0001	
A^5	7.862E+008	1	7.862E+008	405.47	<0.0001	
A^6	3.986E+008	1	3.986E+008	205.57	<0.0001	
RESIDUAL	2.327E+007	12	1.939E+006			
COR TOTAL	5.289E+010	18				

Obser vations:
1. The model F-value of 4544.83 implies the model is significant. There is only a 0.01% chance that a "model F-value" this large could due to noise.
2. Values of " prob>F " less than 0.0500 indicate model terms are significant.
3. In this case A,A^2,A^3,A^4,A^5,A^6 are significant model terms.
4. Values greater than 0.1000 indicate the model terms are not significant.
5. If there are many insignificant model terms(not counting those required to support hierarchy), model reduction may improve the model.

<div align="center">R-Squared Results:</div>

Std.Dev	1392.44	R-squared	0.9996
Mean	14763.58	Adj R-squared	0.9993
C.V.%	9.43	Pred R-squared	0.9945
PRESS	2.901E+008	Adeq precision	184.886

The " pred R-squared"of 0.9945 is in reasonable agreement with the " adj R-squared" of 0.9993.
"adeq precision" measures the signal to noise ratio. A ratio greater than 4 is desirable. The ratio of 184.886 indicates an adequate signal. This model can be used to navigate the design space.
Model equation of respone3:-

R3= 1021.12280+2911.61629*A+102.96956*A^2-4.36850*A^3+0.048767*A^4-2.23065E-004*A^5+3.69608E-007*A^6.

Proceed to diagnostics plots(the next icon in progression). Be sure to look at the:
1.Normal probability plot to the studentized residuals to check for normality of residuals.
2. studentized residuals versus predicted values to check for constant error.
3.Externally studentized residuals to look for outliers, i.e., influential values.
4.Box-Cox plot for power transmissions

If all the model statistics and diagnostic plots are OK, finish up with the model graphs icon.
<div align="center">R3 vs A:A</div>
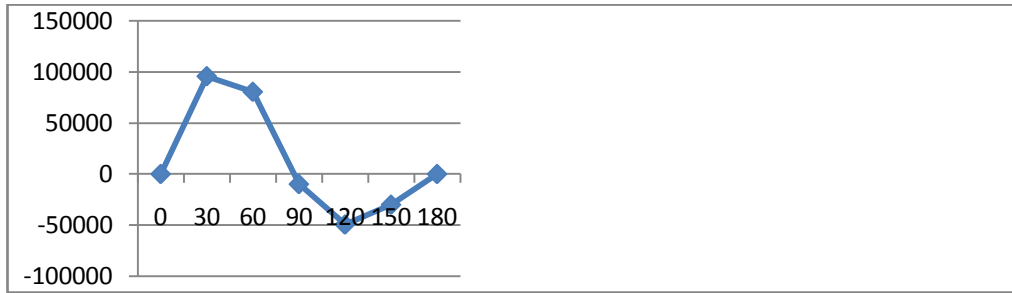
Response 4:

Source	Sum of Squares	df	Mean Square	F value	p-value prob>F	
Model	9.099E+011	6	1.516E+011	3.382E+005	<0.0001	significant
A-A	6.936E+010	1	6.936E+010	1.547E+005	<0.0001	
A^2	9.064E+009	1	9.064E+009	20212.43	<0.0001	
A^3	8.803E+008	1	8.803E+008	1962.89	<0.0001	
A^4	1.170E+009	1	1.170E+009	2608.12	<0.0001	
A^5	1.706E+007	1	1.706E+007	38.04	<0.0001	
A^6	2.608E+008	1	2.608E+008	581.45	<0.0001	
RESIDUAL	5.381E+006	12	4.485E+005			

Obser vations:
1. The model F-value of338154.65 implies the model is significant. There is only a 0.01% chance that a "model F-value" this large could due to noise.
2. Values of " prob>F " less than 0.0500 indicate model terms are significant.
3. In this case A,A^2,A^3,A^4,A^5,A^6 are significant model terms.
4. Values greater than 0.1000 indicate the model terms are not significant.
5. If there are many insignificant model terms(not counting those required to support hierarchy), model reduction may improve the model.

R-Squared Results:

| Std.Dev	669.67	R-squared	1.0000
Mean	-5228.79	Adj R-squared	1.0000
C.V.%	12.81	Pred R-squared	0.9999
PRESS	5.020E+007	Adeq precision	1429.511

The " pred R-squared"of 0.9999 is in reasonable agreement with the " adj R-squared" of 1.0000.
"adeq precision" measures the signal to noise ratio. A ratio greater than 4 is desirable. The ratio of 1429.511 indicates an adequate signal. This model can be used to navigate the design space.
Model equation of respone1:-

> R4=-3.83503E+005+424.21111*A+53.57290*A^2+1.62285*A^3-0.030196*A^4+1.64893E-004*A^5-2.98952E-007*A^6

Proceed to diagnostics plots(the next icon in progression). Be sure to look at the:
1.Normal probability plot to the studentized residuals to check for normality of residuals.
2.Studentized residuals versus predicted values to check for constant error.
3.Externally studentized residuals to look for outliers, i.e., influential values.
4.Box-Cox plot for power transmissions
   If all the model statistics and diagnostic plots are OK, finish up with the model graphs icon.

R4 vs A:A
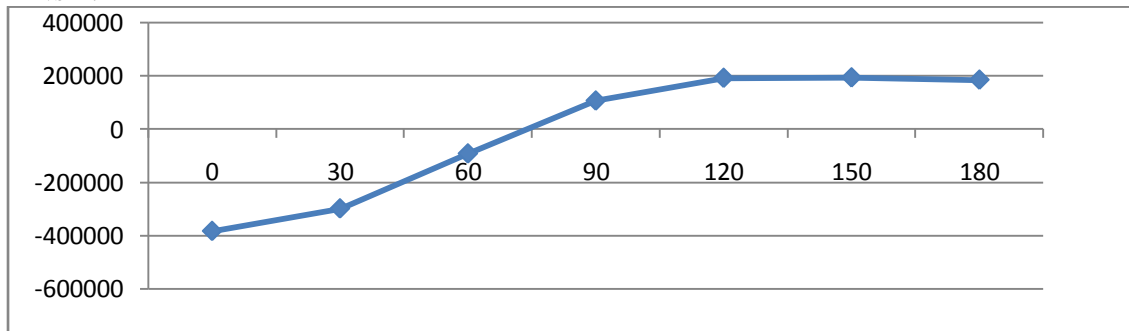
optimization of ga:
The above equations we can substitute MATLAB GA TOOL BOX .from these equations we can get function values.
function y = multi(x)
y(1)=1021.51151+748.60025*x+103.03328*x^2-4.26162*x^3+0.048831*x^4-2.25466e-004*x^5+3.74055e-007*x^6;
y(2)=6.74815e+005-613.87632*x-95.42986*x^2-2.34427*x^3+0.044055*x^4-2.38374e-004*x^5+4.29437e-007*x^6;
y(3)=1021.12280+2911.61629*x+102.96956*x^2-4.36850*x^3+0.048767*x^4-2.23065e-004*x^5+3.69608e-007*x^6;
y(4)=45120.14114-29.45313*x-6.47866*x^2-0.11183*x^3+2.14752e-003*x^4-1.1357e-005*x^5+2.02182e-008*x^6;

index	F1	F2	F3	F4	X1
1.0	1030.010049855302	674808.0295095751	1054.1389002259089	45119.80645920059	0.01133489916838526
2.0	3902.1414740814184	672233.3296832528	10073.316120386307	44980.80958589086	2.854612491185239
3.0	6666.648025134849	669291.1189407776	17217.29226297474	44809.87601864121	4.884400935267964

4.0	13753.518978530545	659385.8033843625	33698.28682758919	44217.60497708893	9.262936403717731
5.0	7712.876010706652	668051.9069602515	19762.042292371087	44736.63501712435	5.580245386297689
6.0	15657.66439670671	656040.1143573725	37970.61293282041	44015.96622404023956	10.374399955568672
7.0	5387.727552646778	670711.0136213038	14006.130279838622	44893.0639376 19634	3.9882269438848876
8.0	17554.694036251265	652354.5283692771	42220.02484493131	43793.58134200431	11.48237864815935
9.0	2487.866063309993	673559.5064616316	5993.01349689371	45054.91452508123	1.6209584805875714
10.0	1030.010049855302	674808.0295095751	1054.1389002259089	45119.80645920059	0.01133489916838526
11.0	9375.89939996724	665932.2898711474	23699.13940182108	44610.49584305366	6.6378561512343115
12.0	1359.1748217631664	674535.6655313867	2281.22082123 92356	45106.39345451995	0.4264668284769069
13.0	11868.025848512572	662388.1884535601	29435.52447059132	44398.18924007252	8.150745196936356
14.0	19231.630991618444	648763.4594354536	45989.58815874372	43576.77213088078	12.471674419307638
15.0	19231.630991618444	648763.4594354536	45989.58815874372	43576.77213088078	12.471674419307638

CONCLUSION
- In this study dynamic reactions investigation was successfully done by using MATLAB software.
- The obtained data have been statistically processed using Response Surface Method.
- The empirical models of output parameters are established and tested through the analysis of variance to validate the adequacy of the models.
- A response surface optimization is attempted using DESIGN EXPERT software for output responses in slider crank mechanism.
- The optimization of slider crank mechanism is done by using GA.

REFERENCES
a. MAT LAB R2009 SOFTWARE HELP .
b. SWILLSON.
c. GA TOOL BOX .
d. DESIGN EXPERT 8.0 HELP.
e. MAT LAB FOR MECHANICAL ENGINEERING TEXTBOOK.
f. WWW.MATLABTUTORIALS.COM.

Selection Problems for Application of Probit, Tobit, Logit & Maximum Likelihood Estimation: A Methodological Issue
BY

[1,]Dr Debasis Patnaik , [2,]Mr Nikunj Sunil Sharma

[1,](Asst Professor, Department Of Economics, BITS Pilani, K K Birla Goa Campus, Goa, India)
[2,](B Tech, Department Of EEE, BITS Pilani, K K Birla Goa Campus, Goa, India)

ABSTRACT:
The application of probabilistic models to economics and finance study poses a problem in the sense of which model is more appropriate. A brief discussion using case studies by authors is undertaken to assess a realistic level of difficulty in the discipline. Then we take recourse to data on women's wages and distribution thereof to assess equity in the system is assess the appropriability of use of a probabilistic model. Assessment of student scorecard is also done to show the relative degree of successful prediction achieved. Stata and SPSS softwares were used for filling in data, testing hypothesis and deriving results to nullify software specificity in result efficiency. Finally a model is built to decide whether an individual decides to pay the requisite amount of taxes or not.

I.   INTRODUCTION

The tobit and probit models are similar in many ways. Each have the same structural model, just different measurement models i.e. how the $y\alpha$ is translated into the observed *y* is different. In the tobit model, we know the value of $y\alpha$ when $y\alpha > 0$, while in the probit model we only know if $y\alpha > 0$. Since there is more information in the tobit model, the estimates of the ¯s should be more efficient. The logistic has slightly flatter tails i.e., the normal or probit curve approaches the axes more quickly than the logistic curve.Qualitatively, Logit and Probit Models give similar results but the estimates of parameters of the two models are not directly comparable. The likelihood function is the joint probability (density) function of observable random variables but it is viewed as the function of the parameters given the realized random variables.

A brief survey of literature on related study of probability models and its applications reveal the following:Carmen Cote and Joseph Farhan (2002) in the paper 'Application of the Two-Stage Bivariate Probit-Tobit Model to Corporate Financing Decisions' used a simulated model aiming to study the factors affecting firms' choice of the form of financing and the size of issue using a two stage *Bivariate Probit – Tobit model. The first stage examines the factors affecting the firms' choice of the form of financing using a Bivariate-Probit model. They used use a two-stage *Bivariate Probit-Tobit model to examine the corporate financing decisions. In this model, managers make three sequential financing decisions that are not necessarily independent. They are: whether to use internal or external source of funding; if external source of funding is the choice, whether to issue debt or equity; make the decision about the size of the debt (equity) issue. The simulation is based on random draws corresponding to 100 years of data for 1,000 firms. The results show that even all firms follow the pecking order behavior, only 85% of the internal and external issuance decisions and less than 70% of the debt and equity issuance decisions are accurately identified. The results show that the correlation coefficients between the Bivariate-Probit equations and those between the Bivariate-Probit and issue size equations (Tobit) are different from zero. This implies that using the Bivariate-Probit model is more appropriate than two independent Probit when studying corporate financing choices. An examination of factors that affect the choice of financing form and the size of issue support the predictions of both trade-off and pecking order theory. Trade-off factors have a significant impact on the debt-equity choice as well as on the size of issue. Firm size and Z-score have a negative impact on the likelihood of using external funding.

1- Carmen Cote And Joseph Farhat , Application of the Two-Stage Bivariate Probit-Tobit Model to Corporate Financing Decisions' in Baker, M., and Wurgler J.(2002), Market Timing and Capital Structure, Journal of Finance 57, 1-32 Henry W. Chappell Jr.(1982) in 'Campaign Contributions and Congressional Voting: A Simultaneous Probit-Tobit Model' aimed to analyze the financial relationships between Interest groups and policymakers empirically. In the analysis of voting on a particular issue, the principal economic agents of concern are congressmen and a single interest group. Congressmen's voting decisions are presumed to be motivated by a desire to be reelected, while the interest group is assumed to allocate campaign funds to various candidates in an attempt to influence the legislative outcome of the issue. A "simultaneous probit-Tobit model" (hence referred to as model SPT) has been hypothesized to explain voting decisions made by congressmen and contribution decisions made by the interest groups. The probit equation is hypothesized to explain votes on the issue. According to the model, a "yes" vote occurs when the unobserved latent variable exceeds a threshold level of zero, and a "no" vote occurs otherwise. This unobserved variable can be interpreted as the candidate's "propensity to vote in favor of the interest group." Interest group contributions are explained by the Tobit equation.

The preceding theoretical discussion provides a basis for the empirical analysis of interest group campaign contributions and roll call voting by members of the U.S. House of Representatives in the 1974-1977 period. Several criteria were used to guide the selection of the seven issues analyzed in the study. First, an effort was made to avoid issues of concern to numerous diverse competing interest groups. Ideally, just one group should be associated with each issue. Issues in regulatory policy often conform to this criterion. It also attempted to select issues for which close votes were recorded in the House, since congressmen may behave differently in their decision- making when voting on issues of certain versus those of doubtful outcomes. Issues for which a congressman must seriously consider the possibility that his vote could influence the ultimate outcome of legislation are preferred. Finally, it was also necessary to choose issues for which an associated interest group made substantial contributions. The seven issues chosen for study include mortgage disclosure requirements for lenders, milk price supports, truck weight limits, tax rebates for oil companies, funding for the B1 bomber, auto emissions controls, and a maritime cargo preference bill. FIML estimates of the simultaneous probit-Tobit model suggest that he effects of campaign Contributions on voting are smaller than single equation probit estimates would indicate. We are generally unable to conclude that contributions have a significant impact on voting decisions; apparently votes are most often decided on the basis of personal ideology or the preferences of constituents. Despite the lack of significance according to model SPT, it would not, however, be appropriate to unambiguously conclude that contributions have no effects on voting. The FIML estimates of the contribution coefficients are not very precise. It is probable that rather poor overall explanatory power in the equations explaining contributions leads to imprecision of these estimates in the voting equation. If better models to explain contributions are developed in the future, this might result in greater precision in estimating the effects of contributions on voting.

2- Henry W. Chappell Jr.(1981) in 'Campaign Contributions and Congressional Voting: A Simultaneous Probit-Tobit Model'*Review of Economics and Statistics*, Volume 64, Issue 1, 1982, pages 77-83. http://www.mitpressjournals.org/loi/rest. *Received for publication December 29, 1980. Revision accepted for publication May 27, 1981. * Universityo f South Carolina.
This paper is based on my Ph.D. dissertationi n economics.

3- Lee C. Adkins in 'An Instrumental Variables Probit Estimator using gretl' aimed at Application of Probit Estimation using gretl (Gnu Regression, Econometrics and Time-series Library) script.And to estimate endogenous probit models Stata 10 was used. Two estimators of this model: a simple `two-step' estimator and a maximum likelihood estimator. Adkins (2008a) compares these estimators to several others in a Monte Carlo study and finds that the two-step estimator performs reasonably well in some circumstances. Gretl script is used to estimate the parameters of an dichotomous choice model that contains endogenous regressors. The routine is simple and yields the same results as the two-step option in the commercially available Stata 10 software. The next step is to duplicate the maximum likelihood estimator, a considerably more challenging undertaking given the multitude of ways the mle can be computed. It should be noted that the only other commercial software that estimates this model via mle is Limdep; [1] Limdep and [2]Stata use different algorithms and yield different results.

Jay Stewart (2009) in 'Tobit or No Tobit?' aim to decide whether to use a Tobit-biased model or not.

*[1]Limdep & [2]Stata are statistical softwares for the estimation of linear and nonlinear regression models and qualitative dependent variable models for cross-section, time-series and panel data. * The GNU General Public License (GNU GPL or simply GPL) is the most widely used <u>free software license</u>. Adkins, Lee C. (2008a), Small sample performance of instrumental variables probit estimators: A monte carlo investigation.Adkins, Lee C. (2008b), `Small sample performance of instrumental variables probit estimators: A monte carlo investigation', Department of Economics, Oklahoma State University, Stillwater OK 74078. available at <u>http://www.learneconometrics.com/pdf/JSM2008.pdf</u>. --Cragg (1971) proposes a double-hurdle model, where the first hurdle is the decision to ever spend money on the good. Since I am restricting my attention to situations where this decision is taken as given, the double-hurdle model reduces to a two-part model. In the first part of the two-part model, a probit is estimated over all observations to determine the probability that individuals purchase the good during the reference period. In the second part, an OLS regression is estimated over the non-zero-value observations. The estimated average probability from the probit is combined with the coefficients from the OLS regression to arrive at unconditional marginal effects.

3-It is published as an IZA Discussion Paper No. 4588 November 2009

Greene Willams[4](2004) in the **'**The behavior of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects'.The general results for the probit and logit models appear to be mimicked by the ordered probit model. Heckman's widely cited result for the probit model appears to be incorrect, however. The differences observed here do not appear to be a function of the mechanism used to generate the exogenous variables. The marginal effects in these binary choice models are overestimated by a factor closer to 50% .A result which has not been considered previously is the incidental parameters effect on estimates of the standard errors of the MLEs. We find that while the coefficients are uniformly overestimated, the asymptotic variances are generally underestimated. This result seems to be general, carrying across a variety of models, independently of whether the biases in the coefficient estimators are towards or away from zero. The ML estimator shows essentially no bias in the coefficient estimators of the tobit model. But the small sample bias appears to show up in the estimate of the disturbance variance. The truncated regression and Weibull [1] models are contradictory, and strongly suggest that the direction of bias in the fixedeffects model is model specific.

5- H. E. RAUCH, F. TUNG AND C. T. STRIEBEL(1965) in 'Maximum Likelihood Estimates of Linear Dynamic Systems'considers the problem of estimating the states of linear dynamic systems in the presence of additive Gaussian noise.

4- Econometrics Journal (2004), volume 7, pp. 98–119. In <u>probability theory</u> and <u>statistics</u>, the **Weibull distribution** is a continuous <u>probability distribution</u>. It is named after <u>Waloddi Weibull</u>, who described it in detail in 1951, although it was first identified by <u>Fréchet (1927)</u> and first applied by <u>Rosin & Rammler (1933)</u> to describe the size distribution of <u>particles</u>.

5-Publication Info: AIAA JOURNAL VOL. 3, NO.-8, AUGUST 1965

Difference equations relating the estimates for the problems of filtering and smoothing are derived as well as a similar set of equations relates the covariance of the errors. The derivation is based on the method of maximum likelihood and depends primarily on the simple manipulation of the probability density functions. The solutions are mechanized on a digital computer. The MLE of the states with continuous observations can be obtained formally from the MLE of the discrete system. The method used here depends primarily on the simple manipulation of the probability density functions and hence leads immediately to recursion equations. The results are also different. The derivation leads directly to a smoothing solution that uses processed data instead of the original measurements.The solution to the discrete version of the filtering and smoothing problem has been derived using the principal of maximum likelihood and simple manipulation of the probabilitydensity function. The filtered estimate is calculated forward point by point as a linear combination of the previous filtered estimate and the current observation. The smoothing solution starts with the filtered estimate at the last point and calculates backward point by point determining the smoothed estimate as a linear combination of the filtered estimate at that point and the smoothed estimate at the previous point. A numerical example has been presented to illustrate the advantage of smoothing in reducing the error in the estimate.

[1]   Wiener, N., The Extrapolation, Interpolation and Smoothing of Stationary Time Series (John Wiley & Sons, Inc., New York,
[2]   1949).
[3]   Parzen, E., "An approach to time series analysis," Ann. Math. Statist. **32,** 951-989 (1961).

METHODOLOGY:

This paper uses Adkins (2008b,a) method to produce a simple routine using the free gretl software. The gretl results are compared to those produced by Stata 10 using data on bank holding companies. The gretl and Stata 10 results are virtually identical. The method of **instrumental variables** (**IV**) is used to estimate causal relationships when controlled experiments are not feasible. Instrumental variable methods allow consistent estimation when the explanatory variables (covariates) are correlated with the error terms of a regression relationship. Time-use surveys collect very detailed information about individuals' activities over a short period of time, typically one day. As a result, a large fraction of observations have values of zero for the time spent in many activities, even for individuals who do the activity on a regular basis. For example, it is safe to assume that all parents do at least some childcare, but a relatively large fraction report no time spent in childcare on their diary day. Tobit seems to be the natural approach. However, it is important to recognize that the zeros in time-use data arise from a mismatch between the reference period of the data (the diary day) and the period of interest, which is typically much longer. Then Tobit doesn't seem appropriate. The bias associated with alternative estimation procedures for estimating the marginal effects of covariates on time use is thus noticed. The bias is often large, and that the extent of the bias increases as the fraction of zero observations increases. It seems likely that one of the main reasons for this poor performance is that the Tobit model assumes that the process that determines whether an individual engages in an activity is the same one that governs how much time is spent in that activity. It adapts the infrequency of purchase model to time-diary data and showing that OLS estimates are unbiased. Next, using simulated data, the bias associated with three procedures that are commonly used to analyze time-diary data – Tobit, the Cragg (1971) two-part model, and OLS under a number of alternative assumptions about the data-generating process. The estimated marginal effects from Tobits are found to be biased and that the extent of the bias varies with the fraction of zero-value observations. The two-part model performs significantly better, but generates biased estimated in certain circumstances. Only OLS generates unbiased estimates in all of the simulations considered here.

METHODOLOGY:

a. **Log likelihood** - This is the log likelihood of the fitted model. It is used in the Likelihood Ratio Chi-Square test of whether all predictors' regression coefficients in the model are simultaneously zero.

b. **Number of obs** - This is the number of observations in the dataset for which all of the response and predictor variables are non-missing.

c. **LR chi2(3)** - This is the Likelihood Ratio (LR) Chi-Square test that at least one of the predictors' regression coefficient is not equal to zero. The number in the parentheses indicates the degrees of freedom of the Chi-Square distribution used to test the LR Chi-Square statistic and is defined by the number of predictors in the model (3).

d. **Prob > chi2** - This is the probability of getting a LR test statistic as extreme as, or more so, than the observed statistic under the null hypothesis; the null hypothesis is that all of the regression coefficients are simultaneously equal to zero. In other words, this is the probability of obtaining this chi-square statistic (22.09) or one more extreme if there is in fact no effect of the predictor variables. This p-value is compared to a specified alpha level, our willingness to accept a type I error, which is typically set at 0.05 or 0.01. The small p-value from the LR test, 0.0001, would lead us to conclude that at least one of the regression coefficients in the model is not equal to zero. The parameter of the chi-square distribution used to test the null hypothesis is defined by the degrees of freedom in the prior line, **chi2(3).**

FIRST MODEL
Effect of Education on Women's Wages

we want to estimate the effect of education on women's wages. The OLS regression for this would be

$yi = xi^- + {}^2i$ ----(1)

where *yi* is the woman's wage and *xi* is her education. The basic selection problem arises in that the sample consists only of women who choose to work and these women may differ in important *unmeasured* ways from women who do not work. For example, women who are smarter may be more likely to enter the labor market. The 'selection equation' for entering the labor market might be:

$Ui = wi° + ui$ ----(2)

where *Ui* represents the utility to woman *i* of entering the labor market and *wi* is a vector of factors known to influence a woman's decision to work such as her education level. *ui* is assumed to be jointly normally distributed with ²*i* and contains any unmeasured characteristics in the selection equation. We don't actually observe *Ui*. All we observe is a dichotomous variable *Zi* with a value of 1 if the woman enters the labor force (*Ui* > 0) and 0 otherwise. So, where does the selection problem actually come from? Well, there are two selection effects.

1. Women with higher levels of education will be more likely to enter the labor force and so we will have a sample of educated women. As Sartori (2003, 114) notes, this non-random aspect of the sample is what is commonly *misunderstood* to be the problem of 'selection bias'. But this on its own does not bias the estimation of the outcome equation in (1).

2. The second selection effect, which is the most important, *is that some uneducated women will go to work*. This is because these women decide that work is worthwhile because they have a high value on some unmeasured variable which is captured in *ui* . In other words, these women get into our sample not because they have high education (they have low values of *wi°*), but because they have large error terms. In contrast, those women who get into our sample because they have high education (large values of *wi°*) will have a more normal range of errors. The problem is that whether or not education (or independent variables of interest in the outcome equation) is correlated with the unmeasured intelligence (our unmeasured variable) in the overall population, these two variables will be correlated in the selected sample. If intelligence does lead to higher wages, then we will underestimate the effect of education on wages because in the selected sample women with little education are unusually smart.
Many dependent variables of interest take only two values (a dichotomous variable), denoting an event or non-event and coded as 1 and 0 respectively. Some

The Logit Model
• When the transformation function *F* is the logistic function, the response probabilities are given by

$$P(y_i = 1 \mid x_i) = \frac{e^{x_i' \beta}}{1 + e^{x_i' \beta}}$$

• And, when the transformation *F* is the cumulative density function (cdf) of the standard normal distribution, the response probabilities are given by

$$P(y_i = 1 \mid x_i) = \Phi(x_i'\beta) = \int_{-\infty}^{x_i'\beta} \Phi(s)\, ds = \int_{-\infty}^{x_i'\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2}\, ds$$

• The Logit and Probit models are almost identical and the choice of the model is arbitrary, although logit model has certain advantages (simplicity and ease of interpretation)

However, the parameters of the two models are scaled differently. The parameter estimates in a logistic regression tend to be 1.6 to 1.8 times higher than they are in a corresponding probit model.

The probit and logit models are estimated by maximum likelihood (ML), assuming independence across observations. The ML estimator of β is consistent and asymptotically normally distributed. However, the

estimation rests on the strong assumption that the latent error term is normally distributed and homoscedastic. If homoscedasticity is violated, no easy solution is found.

In the probit model, use the *Z*-score terminology. For every unit increase in X, the *Z*-score (or the Probit of "success") increases by *b* units. [Or, we can also say that an increase in X changes *Z* by *b* standard deviation units.]

One can convert the z-score to probabilities using the normal table.
• The Tobit model uses all of the information, including info on censoring and provides consistent estimates.
• It is also a nonlinear model and similar to the probit model. It is estimated using maximum likelihood estimation techniques. The likelihood function for the tobit model takes the form:

$$\log L = \sum_{Y_i > 0} -\frac{1}{2}\left[\log(2\pi) + \log \sigma^2 + \frac{(Y_i - \beta X_i)^2}{\sigma^2}\right] + \sum_{Y_i = 0}\log\left[1 - F\left(\frac{\beta X_i}{\sigma}\right)\right]$$

• This is an unusual function, it consists of two terms, the first for non-censored observations (it is the pdf), and the second for censored observations (it is the cdf).

• The estimated tobit coefficients are the marginal effects of a change in *xj* on y*, the unobservable latent variable and can be interpreted in the same way as in a
linear regression model.But such an interpretation may not be useful since we are interested in the effect of *X* on the observable *y* (or change in the censored outcome).
J. Scott Long, 1997 (translated 2002), *Regression Models for Categorical and Limited Dependent Variables*. It can be shown that change in *y* is found by multiplying the coefficient with *Pr(a<y*<b),* that is, the probability of being uncensored. Since this probability is a fraction, the marginal effect is actually attenuated. In the above, *a* and *b* denote lower and upper censoring points. For example, in left censoring, the limits will be: *a* =0, b=+∞.

SECOND MODEL
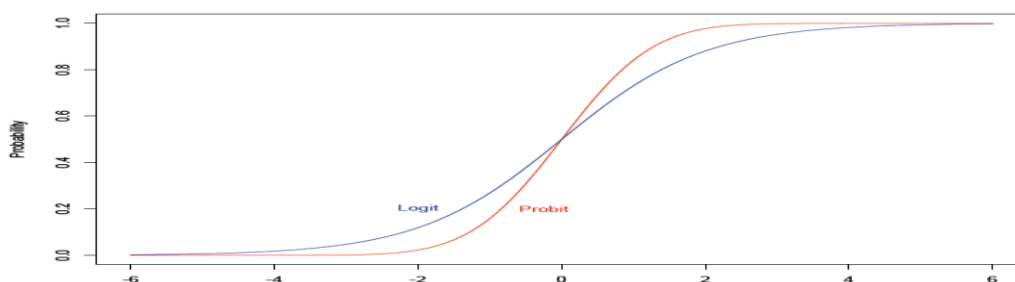
Effect of GRE Scores on Grades in Graduate School
   Suppose that an admissions committee want to know how GRE scores affect the likelihood of success in
graduate school. The problem is that information about success in graduate school (grades) is only available for those students who were admitted. The admissions committee wish to forecast outcomes in the whole pool of applicants but are forced to rely solely on experience with a non-random subset of them. Let's assume that we have the following model. The selection equation for getting admitted might be

Admission Rating = °0 + °1GRE + *ui* ----(3)
Admission = 1 if Admission Rating > 0
     0 if Admission Rating < 0

where ADMISSION RATING is the latent variable measuring the underlying propensity to be admitted, GRE represents a student's GRE score, and Admission is a dichotomous variable indicating whether the student was admitted or not. The outcome equation is

Success = ½ ⁻0 + ⁻1GRE + ²i if Admission=1

Unobserved if Admission = 0

Admitted graduate students are not representative of applicants generally as the admission equation makes clear. There are many college graduates with low grades who attempt to enroll in graduate school; only a few succeed. These exceptions usually owe their success to favorable (unmeasured) personal characteristics other than grades. While many of these personal characteristics will have no affect on their success in graduate school, it seems reasonable to think that some of them probably will. As a result, there will be some students with low grades who make into graduate school because they have large error terms (they have strong personal characteristics). As a result, this low-grade subset of students will perform above the level of other applicants with the same college grades and so they are no longer representative. Now suppose the admissions committee examine graduate student grades to compare the performance of those who entered with low GREs to those who entered with high GREs. The group of students who were admitted because they had strong GREs will be representative of the group of applicants with strong GREs. However, the subset of admitted students with low GREs will not be representative of the group of applicants with low GREs - they will perform better in graduate school (because of their large disturbance terms due to personal characteristics) than applicants with low GREs that were not admitted. Ultimately, it may appear that students with high GREs do not outperform students with low GREs in graduate school. The admissions committee might be tempted to conclude that GREs do not predict success. However, intuition makes it clear that this result does not extend to the applicant pool where students with low GREs would, in general, perform quite poorly had they been admitted. In effect, if a random sample of applicants were admitted to graduate school, GREs would be a good predictor of their success.

THIRD MODEL

Questionnaire :
- On what data is the model being applied.
- Finding the factors affecting the dataset.
- What model to use.
- Calculating the model Coefficients.
- Estimations using either software or hand calculations.
- What has been concluded.

To decide whether an individual decides to pay the requisite amount of taxes or not. And , thus also decide the model to be used for the same.
Below is the dataset for individuals in 17 Latin American countries.

Table 1. Reasons why individuals evade taxes

Why do people not pay their taxes?	Arg	Bol	Braz	Col	Cos	Chi	Ecu	El	Gua	Hon	Mex	Nic	Pan	Par	Per	Uru	Ven	Average
Lack of honesty	17.7	47.0	45.5	31.3	54.0	54.7	53.8	58.5	49.6	53.5	39.2	36.0	49.8	47.3	41.6	20.3	57.5	44.5
Because nationals are quick-witted and sly	14.8	17.6	31.8	17.8	29.2	44.4	47.2	25.8	12.8	28.3	25.4	16.2	32.7	8.9	25.6	30.5	39.9	26.4
They don't see the point in paying taxes	19.7	28.5	25.9	24.4	21.2	30.3	37.8	44.8	15.3	41.3	49.9	30.4	26.9	29.9	21.4	23.4	29.2	29.4
Lack of civic conscience	15.3	35.3	32.0	28.9	24.9	39.5	49.3	40.4	20.2	49.3	38.3	33.4	41.1	37.4	34.2	24.2	40.7	34.4
Because those that evade taxes go unpunished	26.0	23.0	24.3	16.6	19.2	18.1	31.3	36.4	13.4	24.6	36.5	21.0	18.6	19.9	14.6	26.5	22.3	23.1
Because the taxes are ill-spent	26.7	40.4	29.7	40.4	27.8	22.6	45.8	46.4	20.1	35.1	50.3	33.7	27.5	29.9	23.2	25.1	26.6	32.4
Because the taxes are too high	65.6	37.1	50.0	62.8	37.6	32.0	50.8	54.3	24.1	47.2	55.8	57.6	38.5	42.9	50.2	63.7	25.3	46.8
Because there is corruption	32.0	42.4	48.9	48.7	43.7	32.5	59.0	52.5	43.2	44.4	54.6	41.9	40.5	47.0	32.8	41.0	45.7	44.2

Note. Percentage of Individuals that mentioned reasons why people do not pay their taxes.

Using the the two data sets Latinobarometro (data from 1998) and World Values Survey , We apply the standard Probit Model for which Yi* = 1 (for an individual paying his taxes) and Yi* = 0 otherwise for the following set of equations :

Yi* is unobservable but Yi = 0 if Yi* < 0
                              1 if Yi* >= 0

Wherein, P(Yi=1) = P(Yi*>=0) = P(ui >= -B1 -......-Bk*xki)

$$= F(B1 + B2*x2i + ….+Bk*xki)$$

Here, F is the cumulative distributive Function of ui. We are assuming that the probability density function to be symmetric.

All the factors accounted for are listed in the above table.

Now , using Finney's table ,

Thus transforming percentages to probits**.Through Hand Calculations or using Computer Software such as SPSS, SAS , R, or S we can convert the percent responded to probits automatically.

The following figure has been obtained by computing the values in the form of a graph. Its been done using hand calculations.
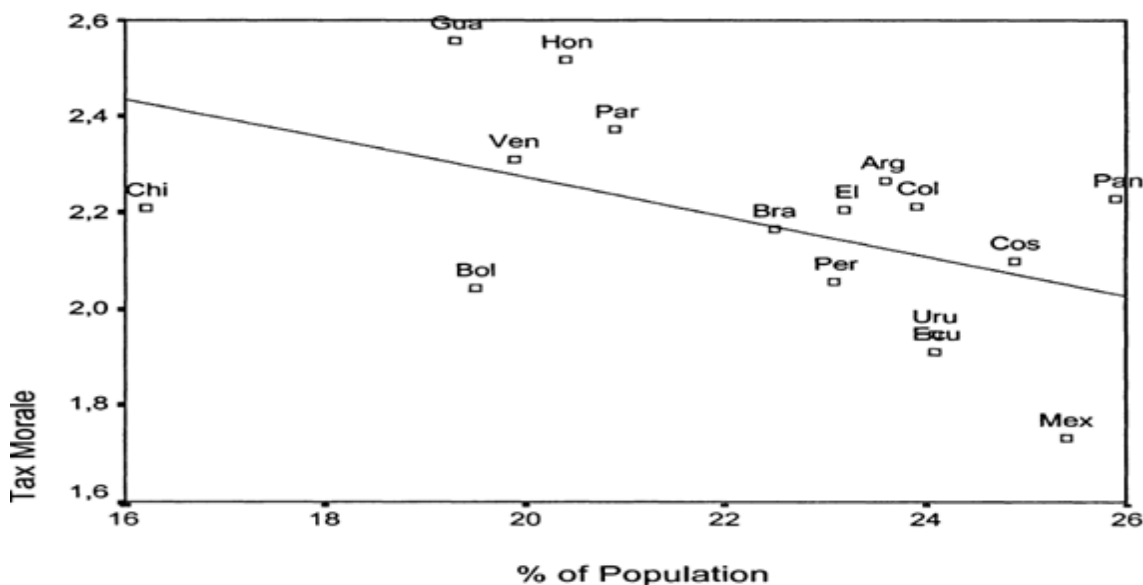
igure 1. Correlation between tax morale and the size of shadow economy

lotes. Arg = Argentina, Bol = Bolivia, Bra = Brazil, Col = Columbia, Cos = Costa ica, Chi = Chile, Ecu = Ecuador, El = El Salvador, Gua = Guatemala, Hon = Honduras, Mex Mexico, Nic = Nicaragua, Pan = Panama, Par = Paraguay, Per = Peru, Uru = Uruguay, Ven Venezuela.

_____

** The Conversion of percentages (w.r.t the various factors affecting the tax morale in various latin American countries) to probits is carried out with the help Finney's Table.

* Finney's Table is shown in Appendix [A] as follows.

Finney's Table :

%	0	1	2	3	4	5	6	7	8	9
0	—	2.67	2.95	3.12	3.25	3.36	3.45	3.52	3.59	3.66
10	3.72	3.77	3.82	3.87	3.92	3.96	4.01	4.05	4.08	4.12
20	4.16	4.19	4.23	4.26	4.29	4.33	4.36	4.39	4.42	4.45
30	4.48	4.50	4.53	4.56	4.59	4.61	4.64	4.67	4.69	4.72
40	4.75	4.77	4.80	4.82	4.85	4.87	4.90	4.92	4.95	4.97
50	5.00	5.03	5.05	5.08	5.10	5.13	5.15	5.18	5.20	5.23
60	5.25	5.28	5.31	5.33	5.36	5.39	5.41	5.44	5.47	5.50
70	5.52	5.55	5.58	5.61	5.64	5.67	5.71	5.74	5.77	5.81
80	5.84	5.88	5.92	5.95	5.99	6.04	6.08	6.13	6.18	6.23
90	6.28	6.34	6.41	6.48	6.55	6.64	6.75	6.88	7.05	7.33
—	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
99	7.33	7.37	7.41	7.46	7.51	7.58	7.65	7.75	7.88	8.09

Finney (1948) has given a table which may be used to test the significance of the deviation from proportionality. As in this case, its been used for converting thefactors affecting tax morale percentages into probits.

The Probit Model
Index function

$$y_i^* = \beta_1 + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + u_i$$

y_i^* is unobservable but $y_i = \begin{cases} 0 \text{ if } y_i^* < 0 \\ 1 \text{ if } y_i^* \geq 0 \end{cases}$

$$P(y_i = 1) = P(y_i^* \geq 0) = P(u_i \geq -\beta_1 - \beta_2 x_{2i} - \ldots - \beta_k x_{ki})$$

$= F(\beta_1 + \beta_2 x_{2i} + \ldots + \beta_k x_{ki})$ where F is the cumulative distribution function

of u_i. We are assuming that the probability density function of u_i is symmetric.

The Logit Model
Its very similar to the probit model.

$$y_i^* = \beta_1 + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + u_i$$

y_i^* is unobservable but $y_i = \begin{cases} 0 \text{ if } y_i^* < 0 \\ 1 \text{ if } y_i^* \geq 0 \end{cases}$

$$P(y_i = 1) = P(y_i^* \geq 0) = P(u_i \geq -\beta_1 - \beta_2 x_{2i} - \ldots - \beta_k x_{ki})$$

$= F(\beta_1 + \beta_2 x_{2i} + \ldots + \beta_k x_{ki})$ where F is the cumulative distribution function

of u_i. We are assuming that the probability density function of u_i is symmetric.

In the probit model we assumed that $u_i \sim N(0, \sigma_u^2)$. In the logit model we assumed that has what is

known as a logistic distribution. The pdf of is given by $f(u_i) = \dfrac{e^{u_i}}{(1 + e^{u_i})^2}$

The model is estimated by MLE.

The Censored Regression (Tobit) Model
The Tobit Model arises when the y variable is limited (or censored) from above or below.

$$y_i^* = \beta_1 + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + u_i$$

y_i^* is unobservable but $y_i = \begin{cases} 0 \text{ if } y_i^* < 0 \\ y_i^* \text{ if } y_i^* \geq 0 \end{cases}$

Binary Probit Regression (in SPSS, use the ordinal regression menu and select probi'
link function. Ignore the test of parallel lines, etc.)

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1645.024			
Final	1166.702	478.322	4	.000

Link function: Probit.

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[work = 0]	2.037	.209	94.664	1	.000	1.626	2.447
Location	age	.035	.004	67.301	1	.000	.026	.043
	education	.058	.011	28.061	1	.000	.037	.080
	children	.447	.029	243.907	1	.000	.391	.503
	[married=0]	-.431	.074	33.618	1	.000	-.577	-.285
	[married=1]	0[a]	.	.	0	.	.	.

Link function: Probit.

a. This parameter is set to zero because it is redundant.

Tobit regression cannot be done in SPSS. Use Stata. Here are the Stata commands.
First, fit simple OLS Regression of the variable lwf (just to check)

. regress lwf age married children education

```
      Source |       SS       df       MS              Number of obs =    2000
-------------+------------------------------           F(  4,  1995) =  134.21
       Model |  937.873188     4  234.468297           Prob > F      =  0.0000
    Residual |  3485.34135  1995  1.74703827           R-squared     =  0.2120
-------------+------------------------------           Adj R-squared =  0.2105
       Total |  4423.21454  1999  2.21271363           Root MSE      =  1.3218

-------------+----------------------------------------------------------------
         lwf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0363624    .003862     9.42   0.000     .0287885    .0439362
     married |   .3188214   .0690834     4.62   0.000     .1833381    .4543046
    children |   .3305009   .0213143    15.51   0.000     .2897004    .3723015
   education |   .0843345   .0102295     8.24   0.000     .0642729    .1043961
       _cons |  -1.077738   .1703218    -6.33   0.000    -1.411765   -.7437105
-------------+----------------------------------------------------------------
```

. tobit lwf age married children education, ll(0)

```
Tobit regression                              Number of obs   =       2000
                                              LR chi2(4)      =     461.85
                                              Prob > chi2     =     0.0000
Log likelihood = -3349.9685                   Pseudo R2       =     0.0645

-------------------------------------------------------------------------
     lwf |     Coef.    Std. Err.      t     P>|t|    [95% Conf. Interval]
---------+---------------------------------------------------------------
     age |   .052157    .0057457     9.08    0.000    .0408888    .0634252
 married |  .4841801    .1035188     4.68    0.000    .2811639    .6871964
children |  .4860021    .0317054    15.33    0.000    .4238229    .5481812
education|  .1149492    .0150913     7.62    0.000    .0853529    .1445454
   _cons | -2.807696    .2632565   -10.67    0.000   -3.323982   -2.291409
---------+---------------------------------------------------------------
  /sigma |  1.872811    .040014                       1.794337    1.951285
-------------------------------------------------------------------------
  Obs. summary:         657  left-censored observations at lwf<=0
                       1343     uncensored observations
                          0  right-censored observations


. mfx compute, predict(pr(0,.))

Marginal effects after tobit
     y  = Pr(lwf>0) (predict, pr(0,.))
        =  .81920975
-------------------------------------------------------------------------
variable |     dy/dx    Std. Err.      z     P>|z|  [    95% C.I.   ]    X
---------+---------------------------------------------------------------
     age |  .0073278    .00083      8.84    0.000    .005703   .008952  36.208
married*|  .0706994    .01576      4.48    0.000    .039803   .101596   .6705
children |  .0682813    .00479     14.26    0.000    .058899   .077663  1.6445
educat~n |  .0161499    .00216      7.48    0.000    .011918   .020382  13.084
-------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1


. mfx compute, predict(e(0,.))

Marginal effects after tobit
     y  = E(lwf|lwf>0) (predict, e(0,.))
        =  2.3102021
-------------------------------------------------------------------------
variable |     dy/dx    Std. Err.      z     P>|z|  [    95% C.I.   ]    X
---------+---------------------------------------------------------------
     age |  .0314922    .00347      9.08    0.000    .024695   .03829   36.208
married*|  .2861047    .05982      4.78    0.000    .168855   .403354   .6705
children |  .2934463    .01908     15.38    0.000    .256041   .330852  1.6445
educat~n |  .0694059    .00912      7.61    0.000    .051531   .087281  13.084
-------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

A researcher is interested in how variables, such as GRE (Graduate Record Exam scores), GPA (grade point average) and prestige of the undergraduate institution, effect admission into graduate school. The response variable, admit/don't admit, is a binary variable.

We have generated hypothetical data, which can be obtained from the URL :
http://www.ats.ucla.edu/stat/stata/dae/probit.htm
This data set has a binary response (outcome, dependent) variable called **admit**. There are three predictor variables: **gre**, **gpa** and **rank**. We will treat the variables **gre** and **gpa** as continuous. The variable **rank** is ordinal, it takes on the values 1 through 4. Institutions with a rank of 1 have the highest prestige, while those with a rank of 4 have the lowest. We will treat **rank** as categorical.
summarize gre gpa

```
Variable |    Obs      Mean   Std. Dev.    Min     Max
-------------+-----------------------------------------------------
     gre |    400     587.7   115.5165     220     800
     gpa |    400    3.3899  .3805668    2.26       4
```

tab rank

```
rank |    Freq.    Percent      Cum.
------------+-----------------------------------
   1 |      61     15.25      15.25
   2 |     151     37.75      53.00
   3 |     121     30.25      83.25
   4 |      67     16.75     100.00
------------+-----------------------------------
Total |     400    100.00
```

tab admit

```
admit |    Freq.    Percent      Cum.
------------+-----------------------------------
   0 |     273     68.25      68.25
   1 |     127     31.75     100.00
------------+-----------------------------------
Total |     400    100.00
```

To run the model in Stata, we first give the response variable (**admit**), followed by our predictors (**gre**, **topnotch** and **gpa**).

tab admit rank

```
|              rank
admit |     1      2      3       4 |   Total
-----------+--------------------------------------------+----------
   0 |     28     97     93      55 |    273
   1 |     33     54     28      12 |    127
-----------+--------------------------------------------+----------
Total |     61    151    121      67 |    400
```

Analysis methods you might consider
Below is a list of some analysis methods you may have encountered. Some of the methods listed are quite reasonable while others have either fallen out of favor or have limitations.
- Probit regression.
- Logistic regression. A logit model will produce results similar probit regression. The choice of probit versus logit depends largely on individual preferences.
- OLS regression. When used with a binary response variable, this model is known as a linear probability model and can be used as a way to describe conditional probabilities. However, the errors (i.e., residuals) from the linear probability model violate the homoskedasticity and normality of errors assumptions of OLS regression, resulting in invalid standard errors and hypothesis tests.

Probit regression
Below we use the **probit** command to estimate a probit regression model. The **i.** before **rank** indicates that **rank** is a factor variable (i.e., categorical variable), and that it should be included in the model as a series of indicator variables. Note that this syntax was introduced in Stata 11.

probit admit gre gpa i.rank

```
Iteration 0:   log likelihood = -249.98826
Iteration 1:   log likelihood = -229.29667
Iteration 2:   log likelihood = -229.20659
Iteration 3:   log likelihood = -229.20658
```

```
Probit regression                    Number of obs  =       400
                                     LR chi2(5)     =     41.56
                                     Prob > chi2    =    0.0000
Log likelihood = -229.20658          Pseudo R2      =    0.0831
```

```
-----------------------------------------------------------------------------
admit |    Coef.    Std. Err.    z   P>|z|   [95% Conf. Interval]
-------------+---------------------------------------------------------------
gre |  .0013756   .0006489    2.12   0.034    .0001038   .0026473
gpa |  .4777302   .1954625    2.44   0.015    .0946308   .8608297
|---------------------------------------------------------------------------
rank |
2 |  -.4153992   .1953769   -2.13   0.033   -.7983308   -.0324675
3 |   -.812138   .2085956   -3.89   0.000   -1.220978   -.4032981
4 |   -.935899   .2456339   -3.81   0.000   -1.417333   -.4544654
|
_cons |  -2.386838   .6740879   -3.54   0.000   -3.708026   -1.065649
-----------------------------------------------------------------------------
```

- In the output above, we first see the iteration log, indicating how quickly the model converged. The log likelihood (-229.20658) can be used in comparisons of nested models, but we won't show an example of that here.
- Also at the top of the output we see that all 400 observations in our data set were used in the analysis (fewer observations would have been used if any of our variables had missing values).
- The likelihood ratio chi-square of 41.56 with a p-value of 0.0001 tells us that our model as a whole is statistically significant, that is, it fits significantly better than a model with no predictors.
- In the table we see the coefficients, their standard errors, the z-statistic, associated p-values, and the 95% confidence interval of the coefficients. Both **gre**, **gpa**, and the three indicator variables for **rank** are statistically significant. The probit regression coefficients give the change in the z-score or probit index for a one unit change in the predictor.
   o For a one unit increase in **gre**, the z-score increases by 0.001.
   o For each one unit increase in **gpa**, the z-score increases by 0.478.
   o The indicator variables for **rank** have a slightly different interpretation. For example, having attended an undergraduate institution of **rank** of 2, versus an institution with a **rank** of 1 (the reference group), decreases the z-score by 0.415.

A test for an overall effect of **rank** using the **test** command can be done. Below we see that the overall effect of **rank** is statistically significant.
test 2.rank 3.rank 4.rank

```
 ( 1)  [admit]2.rank = 0
 ( 2)  [admit]3.rank = 0
 ( 3)  [admit]4.rank = 0

      chi2( 3) =   21.32
    Prob > chi2 =    0.0001
```

We can also test additional hypotheses about the differences in the coefficients for different levels of rank. Below we test that the coefficient for **rank**=2 is equal to the coefficient for **rank**=3.
test 2.rank = 3.rank

(1) [admit]2.rank - [admit]3.rank = 0

    chi2(1) =   5.60
   Prob > chi2 =   0.0179

Model Summary

Parameter Estimates

```
-------------------------------------------------------------------------
admitᵍ|   Coef.ʰ   Std. Err.ⁱ    zʲ  P>|z|ᵏ   [95% Conf. Interval]ˡ
-------------+-----------------------------------------------------------
      gre |  .0015244   .0006382   2.39   0.017   .0002736   .0027752
  topnotch |  .2730334   .1795984   1.52   0.128   -.078973   .6250398
      gpa |  .4009853   .1931077   2.08   0.038   .0225012   .7794694
    _cons | -2.797884   .6475363  -4.32   0.000  -4.067032  -1.528736
-------------------------------------------------------------------------
```

e. **admit** - This is the binary response variable predicted by the model.
gre - The coefficient of **gre** is 0.0015244. This means that an increase in GRE score increases the predicted probability of admission.
topnotch - The coefficient of **topnotch** is 0.2730334. This means attending a top notch institution as an undergraduate increases the predicted probability of admission.
gpa - The coefficient of **gpa** is 0.4009853. This means that an increase in GPA increases the predicted probability of admission.
_cons - The constant term is -2.797884. This means that if all of the predictors (**gre**, **topnotch** and **gpa**) are evaluated at zero, the predicted probability of admission is
$F(-2.797884) = 0.002571929$. So, as expected, the predicted probability of a student with a GRE score of zero and a GPA of zero from a non-topnotch school has an extremely low predicted probability of admission.
To generate values from F in Stata, use the **normal** function. For example,
display normal(0)
will display .5, indicating that $F(0) = .5$ (i.e., half of the area under the standard normal distribution curve falls to the left of zero). The first student in our dataset has a GRE score of 380, a GPA of 3.61, and a topnotch indicator value of 0. We could multiply these values by their corresponding coefficients,
display -2.797884 +(.0015244*380) + (.2730334*0) + (.4009853*3.61)
to determine that the predicted probability of admittance is $F(-0.77105507)$. To find this value, we type
display normal(-0.77105507)
and arrive at a predicted probability of 0.22033715.
f. **Std. Err.** - These are the standard errors of the individual regression coefficients. They are used in both the calculation of the **z** test statistic, superscript j, and the confidence interval of the regression coefficient, superscript l.
g. **z** - The test statistic **z** is the ratio of the **Coef.** to the **Std. Err.** of the respective predictor. The z value follows a standard normal distribution which is used to test against a two-sided alternative hypothesis that the **Coef.** is not equal to zero.
h. **P>|z|** - This is the probability the **z** test statistic (or a more extreme test statistic) would be observed under the null hypothesis that a particular predictor's regression coefficient is zero, given that the rest of the predictors are in the model. For a given alpha level, **P>|z|** determines whether or not the null hypothesis can be rejected. If **P>|z|** is less than alpha, then the null hypothesis can be rejected and the parameter estimate is considered statistically significant at that alpha level.
gre - The **z** test statistic for the predictor **gre** is (0.0015244/0.0006382) = 2.39 with an associated p-value of 0.017. If we set our alpha level to 0.05, we would reject the null hypothesis and conclude that the regression coefficient for **gre** has been found to be statistically different from zero given **topnotch** and **gpa** are in the model.
topnotch - The **z** test statistic for the predictor **topnotch** is (0.2730334/0.1795984) =1.52 with an associated p-value of 0.128. If we set our alpha level to 0.05, we would fail to reject the null hypothesis and conclude that the regression coefficient for **topnotch** has not been found to be statistically different from zero given **gre** and **gpa** are in the model.

gpa - The **z** test statistic for the predictor **gpa** is (0.4009853/0.1931077) = 2.08 with an associated p-value of 0.038. If we set our alpha level to 0.05, we would reject the null hypothesis and conclude that the regression coefficient for **gpa** has been found to be statistically different from zero given **gre** and **topnotch** are in the model.

_cons -The **z** test statistic for the intercept, **_cons,** is (-2.797884/0.6475363) = -4.32 with an associated p-value of < 0.001. With an alpha level of 0.05, we would reject the null hypothesis and conclude that **_cons** has been found to be statistically different from zero given **gre**, **topnotch** and **gpa** are in the model and evaluated at zero.

[95% Conf. Interval] - This is the Confidence Interval (CI) for an individual coefficient given that the other predictors are in the model. For a given predictor with a level of 95% confidence, we'd say that we are 95% confident that the "true" coefficient lies between the lower and upper limit of the interval. It is calculated as the **Coef.** $\pm (z_{\alpha/2})$*(**Std.Err.**), where $z_{\alpha/2}$ is a critical value on the standard normal distribution. The CI is equivalent to the **z** test statistic: if the CI includes zero, we'd fail to reject the null hypothesis that a particular regression coefficient is zero given the other predictors are in the model. An advantage of a CI is that it is illustrative; it provides a range where the "true" parameter may lie.

Illustration (2) :

Example on MLE –

Estimations using R (Statististical Computer Software).

- ► Goal: Try to find the parameter value $\tilde{\beta}$ that makes $E(Y|X,\beta)$ as close as possible to the observed Y

- ► For Bernoulli: Let $p_i = P(Y_i = 1|X_i)$ which implies $P(Y_i = 0|X_i) = 1 - P_i$. The probability of observing Y_i is then

$$P(Y_i|X_i) = P_i^{Y_i}(1 - P_i)^{1-Y_i}$$

- ► Since the observations can be assumed independent events, then

$$P(Y_i|X_i) = \prod_{i=1}^{N} P_i^{Y_i}(1 - P_i)^{1-Y_i}$$

- ► When evaluated, this expression yields a result on the interval $(0, 1)$ that represents the likelihood of observing this sample Y given X if $\hat{\beta}$ were the "true" value

- ► The MLE is denoted as $\tilde{\beta}$ for β that maximizes $L(Y|X,b) = \max L(Y|X,b)$

MLE example: what π for a tossed coin?

	0.5			
Y_i	P^yi	(1-P)^(1-yi)	L	In L
0	1	0.5	0.5	-0.693147
1	0.5	1	0.5	-0.693147
1	0.5	1	0.5	-0.693147
0	1	0.5	0.5	-0.693147
1	0.5	1	0.5	-0.693147
1	0.5	1	0.5	-0.693147
0	1	0.5	0.5	-0.693147
1	0.5	1	0.5	-0.693147
1	0.5	1	0.5	-0.693147
1	0.5	1	0.5	-0.693147
Likelihood			0.0009766	
Log-Likelihood				-6.931472

	0.6			
Y_i	P^yi	(1-P)^(1-yi)	L	In L
0	1	0.4	0.4	-0.916291
1	0.6	1	0.6	-0.510826
1	0.6	1	0.6	-0.510826
0	1	0.4	0.4	-0.916291
1	0.6	1	0.6	-0.510826
1	0.6	1	0.6	-0.510826
0	1	0.4	0.4	-0.916291
1	0.6	1	0.6	-0.510826
1	0.6	1	0.6	-0.510826
1	0.6	1	0.6	-0.510826
Likelihood			0.0017916	
Log-Likelihood				-6.324652

0.7

Y_i		P^yi	(1-P)^(1-yi)	L	In L	
0		1	0.3	0.3	-1.203973	
1		0.7	1	0.7	-0.356675	
1		0.7	1	0.7	-0.356675	
0		1	0.3	0.3	-1.203973	
1		0.7	1	0.7	-0.356675	
1		0.7	1	0.7	-0.356675	
0		1	0.3	0.3	-1.203973	
1		0.7	1	0.7	-0.356675	
1		0.7	1	0.7	-0.356675	
1		0.7	1	0.7	-0.356675	

Likelihood   0.0022236

Log-Likelihood   -6.108643

0.8

Y_i		P^yi	(1-P)^(1-yi)	L	In L	
0		1	0.2	0.2	-1.609438	
1		0.8	1	0.8	-0.223144	
1		0.8	1	0.8	-0.223144	
0		1	0.2	0.2	-1.609438	
1		0.8	1	0.8	-0.223144	
1		0.8	1	0.8	-0.223144	
0		1	0.2	0.2	-1.609438	
1		0.8	1	0.8	-0.223144	
1		0.8	1	0.8	-0.223144	
1		0.8	1	0.8	-0.223144	

Likelihood   0.0016777

Log-Likelihood   -6.390319

MLE example in R

```
> ## MLE example
> y <- c(0,1,1,0,1,1,0,1,1,1)
> coin.mle <- function(y, pi) {
+    lik <- pi^y * (1-pi)^(1-y)
+    loglik <- log(lik)
+    cat("prod L = ", prod(lik), ",  sum ln(L) = ", sum(loglik), "\n")
+    (mle <- list(L=prod(lik), lnL=sum(loglik)))
+ }
> ll <- numeric(9)
> pi <- seq(.1,.9,.1)
> for (i in 1:9) (ll[i] <- coin.mle(y, pi[i])$lnL)
prod L =  7.29e-08 ,   sum ln(L) =  -16.43418
prod L =  6.5536e-06 ,   sum ln(L) =  -11.93550
prod L =  7.50141e-05 ,   sum ln(L) =  -9.497834
prod L =  0.0003538944 ,   sum ln(L) =  -7.946512
prod L =  0.0009765625 ,   sum ln(L) =  -6.931472
prod L =  0.001791590 ,   sum ln(L) =  -6.324652
prod L =  0.002223566 ,   sum ln(L) =  -6.108643
prod L =  0.001677722 ,   sum ln(L) =  -6.390319
prod L =  0.0004782969 ,   sum ln(L) =  -7.645279
> plot(pi, ll, type="b")
```

MLE example in R: plot
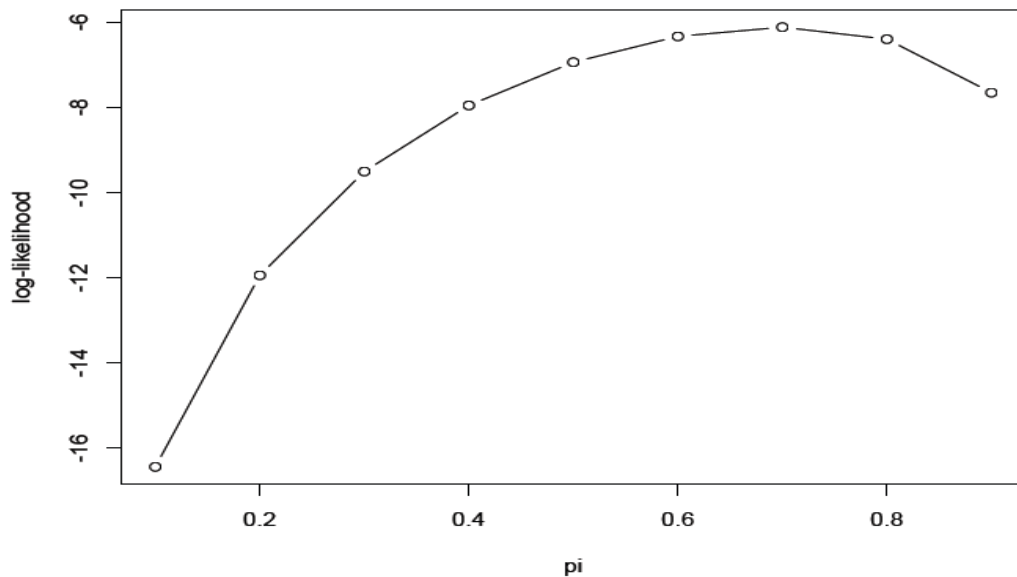
REFERENCES

[1]   Almeida, H. and Campello, M., 2008, Financing Frictions and the Substitution between Internal and External funds, *Journal of Financial and Quantitative Analysis* forthcoming
[2]   Chen, L. and Zhao S., 2004, The modified pecking order theory: New empirical evidence from corporate financing decisions, *Working Paper*, Michigan State University
[3]   Gatchev, V. A., Spindt, P. A., and Tarhan V., 2009, How Do Firms Finance Their Investments? The Relative Importance of Equity Issuance and Debt Contracting Costs, *Journal of Corporate Finance* forthcoming
[4]   Gomes, A. and Phillips G., 2004, Why do public firms issue private and public equity, convertibles and debt?, *Working Paper*, University of Pennsylvania
[5]   Abrams, Burton A., "Legislative Profitsa nd the Economic Theory of RepresentativeV oting:A n Empirical Investigation," Public Choice 31 (Fall 1977), 111-128.
[6]   Alexander, Herbert E., Money in Politics (Washington, D.C.: Public Affairs Press, 1972).
[7]   Chappell, Henry W., "The Economics of Campaign Finance: An Empirical Study," unpublishedP h.D. dissertation, Yale University, 1979.
[8]   McDonald, John and Robert Moffitt (1980) "The Uses of Tobit Analysis." *The Review of*
[9]   *Economics and Statistics* 62(2), May 1980, pp. 318-321.
[10]   Meng, C., and Schmidt P., 1985, On the Cost of Partial Observability in the Bivariate Probit Model, *International Economic Review* 26, 71-85
[11]   Heckman, James J., "Simultaneous Equation Models with Continuous and Discrete Endogenous Variables and Structural Shifts," in Stephen M. Goldfeld and Richard E. Quandt (eds.), Studies in Nonlinear Estiination (Cambridge, MA: Ballinger Publishing Co., 1976).
        URL : http://www.crcnetbase.com/action/showPublications?display=bySubject
        http://www.ats.ucla.edu/stat/stata/dae/probit.htm
        http://www.stata.com/help.cgi?mfx

On The Stability and Accuracy of Some Runge-Kutta Methods of Solving Second Order Ordinary Differential Equations

S.O. Salawu[1], R.A. Kareem[2] and O.T. Arowolo[3]

[1]*Department of Mathematics, University of Ilorin, Ilorin, Nigeria*
[2,3]*Department of Mathematics, Lagos State Polytechnic, Ikorodu, Nigeria*

ABSTRACT.

This paper seeks numerical solutions to second order differential equations of the form $y'' = f(x, y, y')$ with initial value, $y(x_0) = y_0$, $y'(x_0) = y_0'$ using different Runge-Kutta methods of order two. Two cases of Explicit Runge-Kutta method were derived and their stability was determined, this is then implemented. The results were compared with the Euler's method for accuracy.

KEYWORDS. *Euler's Method, Ordinary Differential Equation, Runge-Kutta method, Stability and Taylor series*

I. INTRODUCTION.

Solutions to Differential equation have over the years been a focus to Applied Mathematics. The question then becomes how to find the solutions to those equations. The range of Differential Equations that can be solved by straight forward analytical method is relatively restricted [12]. Even solution in series may not always be satisfactory, either because of the slow convergence of the resulting series or because of the involved manipulation in repeated stages of differentiation [9]. Runge-Kutta methods are among the most popular ODE solvers. They were first studied by Carle Runge and Martin Kutta around 1900. Modern developments are mostly due to John Butcher in the 1960s [4]. The Runge-Kutta method is not restricted to solving only first-order differential equations but Runge-Kutta methods are also used to solve higher order ordinary differential equations or coupled (simultaneous) differential equations [14]. The higher order equations can be solved by considering an equivalent system of first order equations. However, it is also possible to develop direct single steps methods to solve higher order equation.

Definition: We define an explicit Runge-Kutta method with n slopes by the following equations [10].

$$y'' = f(x, y, y') \qquad x \in [x_0, b] \tag{1}$$

$$y_{n+1} = y_n + hy_n' + \sum_{i=1}^{n} r_i k_i \tag{2}$$

$$k_j = \frac{h^2}{2!} f\left(x_n + q_2 h, y_n + q_2 h y_n' + h^2 \sum_{i=1}^{j} q_{ij}, y_n' + \frac{c_{21}}{h} k_1\right) \tag{3}$$

Where $a_{ij}, 1 \le i, j \le n, r_1, r_2 \ldots \ldots r_n$ are arbitrary.

II. DERIVATION OF SECOND ORDER RUNGE-KUTTA METHOD.

To derive the Runge-Kutta methods for second order ordinary differential equation of the form
$y'' = f(x, y, y')$ with the initial condition, $y(x_0) = y_0$, $y'(x_0) = y_0'$.

Let us define

$$k_1 = \frac{h^2}{2!} f(x_n, y_n, y_n') \tag{2.1}$$

$$k_2 = \frac{h^2}{2!} f\left(x_n + q_2 h, y_n + q_2 h y_n' + q_{21} k_1, y_n' + \frac{c_{21}}{h} k_1\right) \tag{2.2}$$

$$y_{n+1} = y_n + h y_n' + r_1 k_1 + r_2 k_2 \tag{2.3}$$

$$y_{n+1}' = y_n' + \frac{1}{h}\left(r_1' k_1 + r_2' k_2\right) \tag{2.4}$$

Where $q_2, q_{21}, b_{21}, r_1, r_2, r_1'$, and r_2' are arbitrary constants to be determined. Using Taylor series expansion form equation (2.3) and (2.4) gives.

$$y_{n+1} = y_n + h y_n' + \frac{h^2}{2!} y_n'' + \frac{h^3}{3!} y_n''' + \frac{h^4}{4!} y_n^{iv} + \ldots \tag{2.5}$$

$$y_{n+1}' = y_n' + h y_n'' + \frac{h^2}{2!} y_n''' + \frac{h^3}{3!} y^{iv}{}_n + \ldots \tag{2.6}$$

$$y_n'' = f(x_n, y(x_n), y'(x_n)) \tag{2.7}$$

$$y_n''' = \left(f_x + y' f_y + f f_{y'}\right)_n \tag{2.8}$$

$$y_n^{iv} = [f_{xx} + y_n'^2 f_{yy} + f^2 f_{y'y'} + 2 y' f f_{xy} + 2 y' f f_{yy'} + 2 f f_{xy'} + f_{y'}\left(f_x + y' f_y + f f_{y'}\right) + f f_y]_n \tag{2.9}$$

We re-write equations (2.7), (2.8) and (2.9) as

$$y_n'' = f_n$$

$$y_n''' = Df_n$$

$$y_n^{iv} = D^2 f_n + f_{y'} Df_n + f_n f_y$$

Where $D = \frac{\partial}{\partial x} + y_n' \frac{\partial}{\partial y} + f_n \frac{\partial}{\partial y'}$

So, equations (2.5) and (2.6) becomes

$$y_{n+1} = y_n + h y_n' + \frac{h^2}{2!} f_n + \frac{h^3}{3!} Df_n + \frac{h^4}{4!}\left(D^2 f_n + f_{y'} Df_n + f_n f_y\right) + \ldots \tag{2.10}$$

$$y_{n+1}' = y_n' + h f_n + \frac{h^2}{2!} Df_n + \frac{h^3}{3!}\left(D^2 f_n + f_{y'} Df_n + f_n f_y\right) + \ldots \tag{2.11}$$

Simplifying k_2, we have

$$\frac{2 k_2}{h^2} = f_n + h\left(q_2 f_x + q_2 y_n' f_y + \frac{1}{2} c_{21} f_n f_{y'}\right) +$$

$$\frac{h^2}{2!}\left(q_2^2 f_{xx} + q_2^2 y_n'^2 f_{yy} + \frac{1}{4} c_{21}^2 f_{y''y'} + 2 q_2^2 y_n' f_{xy} + q_2 c_{21} f_n f_y\right) + O\left(h^3\right)$$

$$k_2 = \frac{h^2}{2} f_n + \frac{h^3}{2} q_2 Df_n + \frac{h^4}{4}\left(q_2^2 Df_n + q_{21} f_n f_{y'}\right) + O\left(h^3\right) \tag{2.12}$$

Where we have used $q_2 = \frac{1}{2} c_{21}$ \hfill (2.13)

The substitution of k_1 and k_2 in equation (2.3) and (2.4) yield

$$y_{n+1} = y_n + h y_n' + \frac{h^2}{2}(r_1 + r_2) f_n + \frac{h^3}{2} q_2 r_2 Df_n + \frac{h^4}{4}\left(r_2 q_2^2 D^2 f_n + r_2 q_{21} f_n f_y\right) + O\left(h^5\right) \tag{2.14}$$

$$y_{n+1}' = y_n' + \frac{h}{2}(r_1' + r_2') f_n + \frac{h^2}{2} q_2 r_2' Df_n + \frac{h^3}{4}\left(r_2' q_2^2 D^2 f_n + r_2' q_{21} f_n f_y\right) + O\left(h^4\right) \tag{2.15}$$

We compare the coefficient of equations (2.14) and (2.15) with equations (2.10) and (2.11) .to obtain.

$$r_1 + r_2 = 1, \quad r_1' + r_2' = 2, \quad q_2 r_2 = \frac{1}{3}, \quad q_2 r_2' = 1 \tag{2.16}$$

The coefficient of h^4 in y_{n+1} and h^3 in y_{n+1}' of equations (2.14) and (2.15) will not match with the corresponding coefficients in equation (2.10) and (2.11) for any choice of q_2, q_{21}, r_2 and r_2'. Thus the local truncation error is $O\left(h^4\right)$ in y and $O\left(h^2\right)$ in y'. A solution of equation (2.12) and (2.16) may be considered for different values of r_1. Here we consider two cases of r_1.

Case I: $r_1 = \frac{1}{2}$

Then, from equation (2.16), we get

$$r_2 = \frac{1}{2}, \quad q_2 = \frac{2}{3} = q_{21}, \quad b_{21} = \frac{4}{3}, \quad r_1' = \frac{1}{2}, \quad r_2' = \frac{3}{2}$$

If the function f is independent of y' then we can construct a Runge-Kutta method in which the local truncation error in y and y' is $O\left(h^4\right)$. Here we obtain,

$$r_1 + r_2 = 1, \quad r_1' + r_2' = 2, \quad q_2 r_2 = \frac{1}{3}, \quad q_2 r_2' = 1, \quad q_2^2 r_2' = \frac{2}{3}, \quad q_{21} r_2' = \frac{2}{3}$$

with the solution

$$r_1 = \frac{1}{2}, \quad q_2 = \frac{2}{3}, \quad q_1 = \frac{1}{2}, \quad r_1' = \frac{1}{2}, \quad r_2' = \frac{3}{2}$$

Thus, the Runge-kutta method for the second order initial value problem

$y'' = f(x, y, y')$ with initial value, $y(x_0) = y_0$, $y'(x_0) = y_0'$

Thus from equation $(2.1), (2.2), (2.3)$ and (2.4) we have

$$k_1 = \frac{h^2}{2!} f(x_n, y_n, y_n'), \quad k_2 = \frac{h^2}{2!} f\left(x_n + \frac{2}{3}h, y_n + \frac{2}{3}hy_n' + \frac{4}{9}k_1, y_n' + \frac{4}{3h}k_1\right) \qquad (2.17)$$

$$y_{n+1} = y_n + hy_n' + \frac{1}{2}(k_1 + k_2), \quad y_{n+1}' = y_n' + \frac{1}{2h}(k_1 + 3k_2) \qquad (2.18)$$

Case II: $r_1 = 0$

From equation (2.16), we have

$$r_2 = 1, q_2 = \frac{1}{3}, r_2' = 1, r_1' = 1, b_{21} = \frac{2}{3}, q_{21} = \frac{1}{9}$$

Therefore,

$$k_1 = \frac{h^2}{2!} f(x_n, y_n, y_n'), \quad k_2 = \frac{h^2}{2!} f\left(x_n + \frac{1}{3}h, y_n + \frac{1}{3}hy_n' + \frac{1}{9}k_1, y_n' + \frac{2}{3h}k_1\right) \qquad (2.19)$$

$$y_{n+1} = y_n + hy_n' + k_2, \quad y_{n+1}' = y_n' + \frac{1}{h}(k_1 + k_2) \qquad (2.20)$$

The two methods derived are explicit second order Runge-Kutta.

III.   STABILITY ANALYSIS.
     While numerically solving an initial value problem for ordinary differential equations, an error is introduced at each integration step due to the inaccuracy of the formula. Even when the local error at each step is small, the total error may become large due to accumulation and amplification of these local errors. This growth phenomenon is called numerical instability [11].
We shall discuss the stability of the Runge-Kutta method in (2.18) and (2.20)
Let us consider the differential equation

$$y'' = \alpha y \qquad (3.1)$$

Subject to initial condition

$$y(x_0) = y_0, \quad y'(x_0) = y_0', \quad x \in [x_0, b]$$

Where α is a real number

We shall consider the case $\alpha = -k^2$ and $\alpha = k^2$ while $\alpha = 0$ give a trival solution. Recall that the second order ordinary differential equation is of the form $y'' = f(x, y, y')$

Then, $k_1 = \dfrac{h^2}{2} y''$ from equation (3.1)

Substitute this (3.1) in equation (2.17), we have

$$k_1 = \frac{h^2}{2}\alpha y_n, \qquad k_2 = \left(\frac{h^2}{2}\alpha + \frac{h^4}{9}\alpha^2\right)y_n + \frac{h^3}{3}\alpha y_n' \tag{3.2}$$

Substituting the equation (3.2) into equation (2.18), we have $y_{n+1} = \left(1 + \dfrac{h^2}{2}\alpha + \dfrac{h^4}{18}\alpha^2\right)y_n + \left(h + \dfrac{h^3}{6}\alpha\right)y_n',$

$$y_{n+1}' = \left(\alpha h + \frac{h^3}{6}\alpha^2\right)y_n + \left(1 + \frac{h^2}{2}\alpha\right)y_n' \tag{3.3}$$

Let us now consider the case when $\alpha = -k^2$. The solution in this case is oscillating.

We, therefore consider the eigenvalues of the matrix $\begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}$ which is given by

$$\lambda = \frac{(q_{11} + q_{22}) \pm \left[(q_{11} - q_{22})^2 + 4q_{12}q_{21}\right]^{1/2}}{2}$$

If $\lambda = \lambda_1, \lambda_2$, then

$$\lambda_1, \lambda_2 = \frac{(q_{11} + q_{22}) \pm \left[(q_{11} - q_{22})^2 + 4q_{12}q_{21}\right]^{1/2}}{2} \tag{3.4}$$

To determine the value of q_{11}, q_{12}, q_{21}, and q_{22}, we consider

$$\begin{bmatrix} y_{n+1} \\ y_{n+1}' \end{bmatrix} = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}\begin{bmatrix} y_n \\ y_n' \end{bmatrix} \tag{3.5}$$

We can re-write equation (3.5) as

$$y_{n+1} = q_{11}y_n + q_{12}y_n'$$

$$y_{n+1}' = q_{21}y_n + q_{22}y_n' \tag{3.6}$$

when $\alpha = -k^2$ we have

$$q_{11} = 1 - \frac{h^2}{2}k^2 + \frac{h^4}{18}k^4, q_{12} = h - \frac{h^3}{6}k^2, q_{21} = -hk^2 + \frac{h^3}{6}k^4, q_{22} = 1 - \frac{h^2}{2}k^2 \qquad (3.7)$$

Substituting equation (3.7) in (3.4), we have

$$\lambda_1, \lambda_2 = \frac{1}{2}\left[\left(2 - h^2k^2 + \frac{h^4k^4}{18}\right) \pm \left[\left(\frac{hk}{18}\right)^2 \left(h^6k^6 - 36h^4k^4 + 432h^2k^2 - 1296\right)\right]^{1/2}\right]$$

$$\lambda_1, \lambda_2 = \frac{1}{2}\left[\left(2 - h^2k^2 + \frac{h^4k^4}{18}\right) \pm \left[\left(\frac{hk}{18}\right)^2 \left(h^2k^2 - 4.44044737\right)\left(h^4k^4 - 2\eta h^2k^2 + \eta^2 + \varphi^2\right)\right]^{1/2}\right]$$

Where $\eta = 15.779763$ and $\varphi = 6.5467418$

Computing λ_1 and λ_2 as functions of h^2k^2, we find that the roots have unit modulus $0 \le h^2k^2 \le 4.44$.

Thus, the stability interval of the Runge-Kutta method (2.18) is $0 \le h^2k^2 \le 4.44$. which also show that is of order 2.

For $\alpha = k^2$, the solutions of (3.1) are exponential in nature and can be written in the matrix form as

$$\begin{bmatrix} y(x_n) \\ y'(x_n) \end{bmatrix} = e^{(x-x_0)b} \begin{bmatrix} y_0 \\ y'_0 \end{bmatrix}$$

And $b = \begin{bmatrix} 0 & 1 \\ k^2 & 0 \end{bmatrix}$ for point $x = x_0 + nh = x_n$, this solution becomes

$$\begin{bmatrix} y(x_n) \\ y'(x_n) \end{bmatrix} = \begin{bmatrix} 1 + \frac{h^2k^2}{2} + \frac{h^4k^4}{24} + \ldots & h + \frac{h^3k^2}{6} + \ldots \\ hk^2 + \frac{h^3k^4}{6} + \ldots & 1 + \frac{h^2k^2}{2} + \frac{h^4k^4}{24} + \ldots \end{bmatrix} \begin{bmatrix} y_0 \\ y'_0 \end{bmatrix}$$

And then,

$$\lim_{h \to 0} \frac{1}{h^4}\left[e^{bh} - \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}\right] = \begin{bmatrix} -\frac{1}{72}k^4 & 0 \\ 0 & \frac{1}{24}k^4 \end{bmatrix}$$

The relative error for the method under discussion in case of a large number of integration interval (large x, small h) are to be considered.

The maximum eigenvalue of the matrix.

$$
\begin{bmatrix}
1 + \dfrac{h^2 k^2}{2} + \dfrac{h^4 k^4}{18} & h + \dfrac{h^3 k^2}{6} \\[3mm]
hk^2 + \dfrac{h^3 k^4}{6} & 1 + \dfrac{h^2 k^2}{2}
\end{bmatrix}
$$

Is obtained as

$$
\lambda = \frac{1}{2}\left[(q_{11} + q_{22}) + \left[(q_{11} - q_{22})^2 + 4 q_{12} q_{21} \right]^{\frac{1}{2}} \right]
$$

$$
\lambda = 1 + hk + \frac{1}{2} h^2 k^2 + \frac{1}{6} h^3 k^3 + \frac{1}{36} h^4 k^4 + \ldots
$$

And therefore the relative error is given by

$$
F_\infty \approx \frac{hk - \log \lambda}{h} = \frac{\log\left(e^{hk}\right) - \log \lambda}{h} = \frac{1}{h}\log\left[1 + \frac{e^{hk} - \lambda}{\lambda} \right] \approx \frac{1}{72} h^3 k^3
$$

We shall now consider the stability of the Runge-Kutta method (2.20) using differential equation (3.1). Equation (2.19) becomes

$$
k_1 = \frac{h^2}{2}\alpha y_n, \qquad k_2 = \left(\frac{h^2}{2}\alpha + \frac{h^4}{36}\alpha^2 \right) y_n + \frac{h^3}{6}\alpha y_n' \tag{3.8}
$$

Substituting the equation (3.9) into equation (2.21), we have

$$
y_{n+1} = \left(1 + \frac{h^2}{2}\alpha + \frac{h^4}{36}\alpha^2 \right) y_n + \left(h + \frac{h^3}{6}\alpha \right) y_n', \quad y_{n+1}' = \left(\alpha h + \frac{h^3}{36}\alpha^2 \right) y_n + \left(1 + \frac{h^2}{6}\alpha \right) y_n' \tag{3.9}
$$

We now consider the case when $\alpha = -k^2$. The eigenvalues is given by

$$
\lambda_1, \lambda_2 = \frac{1}{2}\left[\left(2 - \frac{h^2 k^2}{3} + \frac{h^4 k^4}{36} \right) \pm \left[\left(\frac{hk}{36} \right)^2 \left(h^6 k^6 - 48 h^4 k^4 + 1152\, h^2 k^2 - 5184 \right) \right]^{\frac{1}{2}} \right]
$$

$$
\lambda_1, \lambda_2 = \frac{1}{2}\left[\left(2 - \frac{h^2 k^2}{3} + \frac{h^4 k^4}{36} \right) \pm \left[\left(\frac{hk}{36} \right)^2 \left(h^2 k^2 - 5.688510232 \right)\left(h^4 k^4 - 2\eta h^2 k^2 + \eta^2 + \varphi^2 \right) \right]^{\frac{1}{2}} \right]
$$

Where $\eta = 21.15574488$ and $\varphi = 26.23461232$

Computing λ_1 and λ_2 as functions of $h^2 k^2$, the roots have unit modulus $0 \leq h^2 k^2 \leq 5.69$ Thus, the stability interval of the Runge-Kutta method (2.21) is $0 \leq h^2 k^2 \leq 5.69$ and is of order 2.

For $\alpha = k^2$, the solutions of (3.1) are exponential in nature as above in case I.

The maximum eigenvalue of the matrix.

$$
\begin{bmatrix}
1 + \dfrac{h^2 k^2}{2} + \dfrac{h^4 k^4}{36} & h + \dfrac{h^3 k^2}{6} \\[3mm]
hk^2 + \dfrac{h^3 k^4}{36} & 1 + \dfrac{h^2 k^2}{6}
\end{bmatrix}
$$

Is obtained as

$$
\lambda = \frac{1}{2}\left[(q_{11} + q_{22}) + \left[(q_{11} - q_{22})^2 + 4 q_{12} q_{21} \right]^{\frac{1}{2}} \right]
$$

$$
\lambda = 1 + hk + \frac{1}{3}h^2 k^2 + \frac{1}{6}h^3 k^3 + \frac{1}{72}h^4 k^4 + \ldots
$$

The relative error is given by

$$
F_\infty \approx \frac{hk - \log \lambda}{h} = \frac{\log(e^{hk}) - \log \lambda}{h} = \frac{1}{h}\log\left[1 + \frac{e^{hk} - \lambda}{\lambda} \right] \approx \frac{1}{108}h^3 k^3
$$

IV. IMPLEMENTATION OF THE METHOD.

Consider the initial value problem [13]

$y'' - y' = x$ Subjected to initial condition $y(0) = 1$, $y'(0) = 2$, $h = 0.1$ \hfill (4.1)

The complementary solution is $y_c = A + Be^x$ and the particular solution gives $y_p = -\frac{1}{2}x^2 - x$ Using the initial condition, we have $A = -2$ and $B = 3$

The general solution become

$$
y = -2 + 3e^x - \frac{1}{2}x^2 - x
$$

Numerical solutions are preferred to the derived cases I and case II of explicit second order Runge-Kutta method of the IVP in (4.1), obtaining numerical solutions for values of x up to and including $x = 1$ with a step size of 0.1 as found in Table 1 and Table 2.

Table1: Solutions of case I second ordered Runge-Kutta methods with $h = 0.1$

X	Numerical solution y(x)	Analytical solution Y(x)	Absolute Error
0	1.000000000	1.000000000	0.000000000
0.1	1.210500000	1.210512754	1.288×10^{-5}
0.2	1.444127500	1.444208274	8.08×10^{-5}
0.3	1.704360888	1.704576424	2.16×10^{-4}
0.4	1.995043782	1.995474094	4.30×10^{-4}
0.5	2.320423378	2.321163813	7.40×10^{-4}
0.6	2.685192833	2.686356400	1.16×10^{-3}
0.7	3.094538080	3.096258121	1.72×10^{-3}
0.8	3.554189578	3.556622784	2.43×10^{-3}

| 0.9 | 4.070479483 | 4.073809330 | 3.33×10^{-3} |
| 1.0 | 4.650404829 | 4.654845484 | 4.44×10^{-3} |

Figure 1:

Table2: Solutions of case II second ordered Runge-Kutta methods with $h = 0.1$

X	Numerical solution y(x)	Analytical solution Y(x)	Absolute Error
0	1.000000000	1.000000000	0.000000000
0.1	1.210500000	1.210512754	1.28×10^{-5}
0.2	1.443075833	1.444208274	1.13×10^{-3}
0.3	1.700988542	1.704576424	3.59×10^{-3}
0.4	1.987830710	1.995474094	7.64×10^{-3}
0.5	2.307560166	2.321163813	1.36×10^{-2}
0.6	2.664537116	2.686356400	2.18×10^{-2}
0.7	3.063565056	3.096258121	3.27×10^{-2}
0.8	3.509935837	3.556622784	4.67×10^{-2}
0.9	4.009479314	4.073809330	6.43×10^{-2}
1.0	4.568618045	4.654845484	8.62×10^{-2}

Figure 2:

Euler's method

Euler's method is one of many methods for generating numerical solutions to differential equations. Besides, most of the other methods that might be discussed are refinements of Euler's method. This method is implemented and compared its accuracy and the error with the method in section 4.
The Euler's method generalized in the form [11].

$$y_{n+1} = y_n + hf\left(x_n, y_n, y_n'\right) \qquad (5.1)$$

$$y_{n+1}' = y_n' + hf\left(x_n, y_n, y_n'\right)$$

(5.2) Consider the initial value problem in equation (4.1)

$$y'' - y' = x$$

Subjected to initial condition

$$y(0) = 1, y'(0) = 2, h = 0.1$$

Table3: Solutions of Euler's methods with $h = 0.1$

X	Numerical solution y(x)	Analytical solution Y(x)	Absolute Error
0	1.000000000	1.000000000	0.000000000
0.1	1.200000000	1.210512754	1.05×10^{-2}
0.2	1.430000000	1.444208274	1.42×10^{-2}
0.3	1.693000000	1.704576424	1.16×10^{-2}
0.4	1.992300000	1.995474094	7.25×10^{-3}
0.5	2.331530000	2.321163813	1.04×10^{-2}
0.6	2.714680000	2.686356400	2.83×10^{-2}
0.7	3.146145000	3.096258121	4.99×10^{-2}
0.8	3.630756500	3.556622784	7.41×10^{-2}
0.9	4.173829150	4.073809330	1.00×10^{-1}

1.0	4.781209065	4.654845484	1.26×10^{-1}

Figure 3:

V.   CONCLUSION.

Investigation carried out on some explicit second ordered Runge-Kutta method in this paper has shown that the stability interval of the Runge-Kutta method in case I is $0 \leq h^2 k^2 \leq 4.44$ and the stability interval of the Runge-Kutta method in case II is $0 \leq h^2 k^2 \leq 5.69$. It is clear that case I is more stable than case II of the derived Runge-Kutta methods. The two methods are shown to be accurate, efficient and general in application for sufficiently solution of $y(x)$. The result obtained in the present work demonstrate the effectiveness and superiority for the solution of second order ordinary differential equation which gave a very high accuracy when compared with exact solution.

REFERENCE.

[1]   Autar Kaw, Runge-Kutta Second Order Method for Ordinary Differential Equations. University of South Florida; Holistic Numerical methods Institute, 2013.

[2]   Buck R.C. and Buch E.F., Introduction to Differential Equation; Honghton Mifflin Company, 1976

[3]   Butcher J.C., Numerical Methods for Ordinary Differential Equations; John Wiley and sons, New York, 2003.

[4]   [Butcher John, Runge-Kutta Methods for Ordinary Differential Equations; COE Workshop on Numerical Analysis Kyushu University, 2005.

[5]   Christopher P. Grant, Theory of Ordinary Differential Equations; Brigham Young University.

[6]   Corliss G. and Kirlinger, On Implicit Taylor Series methods for Stiff Ordinary Differential Equation; Centre for Research on Parallel Computation, Rice University Houston, 1991.

[7]   Curtis F.G. and Patrick O.W., Applied Numerical Analysis; Addison Wesley Publishing Company, 1989.

[8]   Greenspan D., Theory and Solutions of Ordinary Differential Equations; The MacMillan Company, New York, 1960.

[9]   Iserles A., Numerical Analysis for Ordinary Differential Equations; Cambridge University Press Cambridge, 1996.

[10]   Jain M.K., Numerical Solution of Differential Equations, Wiley Eastern Limited.

[11]   Joe D. Hoffman, Numerical Methods for Engineers and Scientists; marcel Dekker Inc, 2001.

[12]   Recktenwald G., Numerical Integration of ODEs for Initial Value Problems. Department of Mechanical engineering, Portland State University. 2006.

[13]   Okunnuga S.A. and Akanbi M.A., Computational Mathematics; A First Course. Wim Publication Lagos Nigeria, 2003.

[14]   Patil P.B. and Verma U.P. Numerical Computational Methods, Narosa Publishing House, New Delhi.

FPGA Modeling Of Neuron for Future Artificial Intelligence Applications

S. Sai Sree Andal[1] (M.Tech), N. Aravind[2], Asst. Prof.

[1, 2,] *Department of ECE, Teegala Krishna Reddy Engineering College/JNTU, India*

ABSTRACT:

An Artificial Neural Network, often just called a neural network, is a mathematical model inspired by biological neural networks. A neural network consists of an interconnected group of artificial Neurons. An artificial neuron is a mathematical function conceived as a crude model, or abstraction of biological neurons. This project describes a system realization of translating data from electrochemical sensor for neuron to process on FPGA. The structure of a neuron is split into various sub blocks and these blocks will be implemented individually first and then they are integrated to form the entire neuron. This project will be implemented in three stages. First we have to convert the analog signal coming from the analog circuitry using an ADC (analog to digital converter). In this project the 12 bit ADC chip will be used to convert analog signal from 4 channel analog circuitry to digital. The next module is the design of mathematical operation. This includes issues relating to data structure, design of Multiplier Accumulator (MAC) and activation function implementation. The final module is displaying the result from the data that have been accumulated by the neuron. The proposed architecture is simulated using Modelsim and synthesized using Xilinx ISE and it will be implemented on FPGA board for hardware implementation and testing. The Xilinx Chip scope tool will be used to test the FPGA inside results while the logic running on FPGA.

KEYWORDS: *Synaptic weights, Activation function, Electrochemical sensors, Bias*

1. INTRODUCTION

Electrochemical sensors are often used to determine concentrations of various analytes in testing samples such as fluids and dissolved solid materials.   Electrochemical sensors are frequently used in occupational safety, medical engineering, process measuring engineering, environmental analysis.ANN   is known to be able to improve electrochemical sensor this signal interpretation. In general, hardware realization requires a good compromise between accuracy and complexity of the processing units to allow a low cost effective device. In our project we are going to implement an artificial neuron with four inputs. The artificial neuron which we are implementing in our project is a prototype of the biological neuron. Basically a neuron consists of N inputs coming from dendrites get multiplied by the synaptic weights and then they are processed by soma .Depending on the strength of the input signals the neuron gets fired. Similarly we are going to implement an artificial neuron with 4 inputs. The Inputs are provided by an analog circuitry which has four channels. This analog circuitry can provide the Analog voltages for 4 channels .This acts as input source for our artificial neuron. Since our FPGA cannot deal with the analog voltages, we convert the analog input into digital output. For this purpose we are going to use 2 ADC modules in our project. The digital platform is Field Programmable Grid Array (FPGA).The approach for this project can be represented in block diagram as shown in the Fig.1. The key issue in designing this system is modular design for reconfigurability. The first issue is to convert the signal from an analog to a digital form, by sampling it using an analog-to-digital converter (ADC), which turns the analog signal into a stream of numbers.
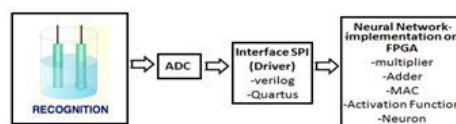

Fig.1.General flow of the project system linking applied chemical sensor to digital processing

The next module is the design of mathematical operation. This includes issues relating to data structure, design of Multiplier Accumulator (MAC) and activation function implementation. The final module is displaying the result from the data that have been accumulating by the neuron. Chipscope tool is used to view the Results after dumping the bit file into the FPGA.

II.DESIGN AND METHODOLOGYS

This section presents the design of the sub modules in implementing Fig 1. This covers the interfacing issues such as analog to digital implementation, data structure and the neuron architecture topology.

2.1.Analog to Digital Interfacing

In this project the 12 bit ADC 7476 were used to convert analog signal from electrochemical sensor to digital. It is a successive approximation 12-bit A/D converter with onboard sample and hold circuitry. As we know that the analog to digital converter is used to convert the analog signal into the digital samples. And we are using the 12 bit A/D converter means at the ADC output we get the 12 bit sample values of the giving a message signal (Modulating signal) through ADC. It is 2-channel ADC means we can give 2 inputs at a time and here we are connecting the input to the ADC from the Function Generator by giving the frequency levels and selecting a wave(Sine or Sawtooth, Square..etc) and constant voltage levels.. Communication with the device is done using a simple serial interface compatible with the SPI protocol. SPI is an interface that allows one chip to communicate with one or more other chips and in this case is ADC 7476 with the FPGA-Spartan 3E board.

The SPI algorithm is required to be implemented in hardware description language (HDL) on FPGA.Fig.2. demonstrates SPI interfacing that allows one chip to communicate with one or more other chips. As shown in the figure above the wires are called SCK, MOSI, MISO and SSEL, and one of the chip is called the SPI master, while the other the SPI slave. A clock is generated by the master, and one bit of data is transferred each time the clock toggles. Data is serialized before being transmitted, so that it fits on a single wire. There are two wires for data, one for each direction. The master and slave know beforehand the details of the communication (bit order, length of data words exchanged, etc...). The master is the one who initiates communication. Because SPI is synchronous and full-duplex, every time the clock toggles, two bits are actually transmitted (one in each direction). In term of performance, SPI can easily achieve a few Mbps (mega-bits-per-seconds) [4]. For this module, the approach taken is hardware implementation of existing technique, tailored to 12-bit environment.
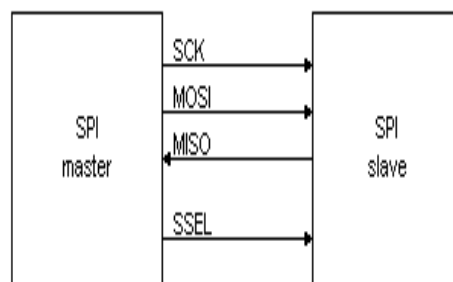

Fig.2.SPI Interfacing applied to signal

2.2. Digital Design: Data Structure and Modules for Neuron

In this section, there are 2 major parts: data structure and digital modules for neuron design on FPGA. For design tools, Modelsim to simulate the design at multiple stages throughout the design process and Quartus to program the board are used. Generally, a data structure is a particular way of storing and organizing data in a computer so that it can be used efficiently. Data structures are generally based on the ability of a computer/chip to fetch and store data at any place in its memory, specified by an address that can be manipulated by the program. For this project, the data computed from ADC will be converted into fixed-point number representation.

Fixed-point DSPs use 2's complement fixed-point numbers in different Q formats. Among the major issues in data structure is the conversion technique of fixed-point number from a Q format to an integer value so that it can be stored in memory and recognized by simulator. It is also required to keep track of the position of the binary point when manipulating fixed- point numbers in writing verilog codes The DSP (Digital Signal Processing) flows throughout the conversion to Q format representation are shown in the Fig. 3. As shown in the flowchart, a fractional number is converted to an integer value that can be recognized by a DSP assembler using the Q15 format .The number is first normalize then scaled down by 2 to the appropriate value that can be accommodated by the bits number. Finally, the value will be rounded (truncated) to integer value and be represented in binary number. A neuron can be viewed as processing data in three steps; the weighting of its input values, the summation of them all and their filtering by a activation function.

The Neuron can be expressed by the following equation:

$$y_j = f\left(\sum_i w_{ij} x_i - \theta_j\right) \tag{1}$$

where y is the output of the neuron, w is the synaptic weight, x is the input and θ is the bias. The subscript i denotes the preceding neuron and j the neuron considered. The neuron computes the product of its inputs, with the corresponding synaptic weights, and then the results are added. The result is presented to a comparison unit designed to represent an appropriate activation function such as linear, sigmoid or hyperbolic tangent. The equation is shown in block diagrams in Fig.4. For the weighted inputs to be calculated in parallel using conventional design techniques, a large number of multiplier units would be required. To avoid this, multiplier/Accumulator architecture has been selected. It takes the input serially, multiplies them with the corresponding weight and accumulates their sum in a register.
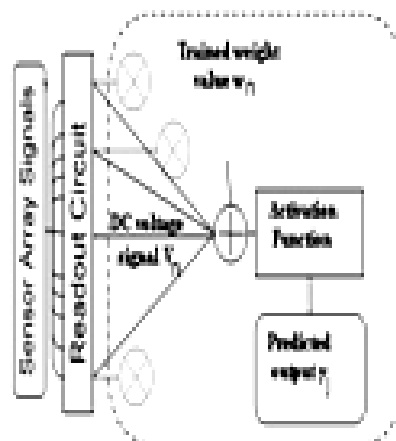

fig.4., Structure of neuron

The block diagram and flow of the hardware implementation is shown in Fig. 5 and 6. The accumulator unit is composed of a bit-serial adder and 16 bit register. The design of multiplier accumulator consists of adder and multiplier. MAC are frequently used in general computing and are especially critical to performance of digital signal processing applications. The MAC typically operate on a digital, and usually binary, multiplier quantity and a corresponding digital multiplicand quantity and generate a binary product. The design of multiplier accumulator proposed in this project consists of adder and multiplier that can accommodate or handle 4 channel of input (array of sensor).
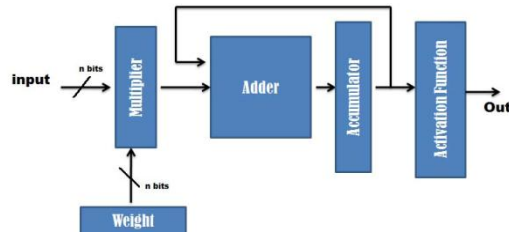
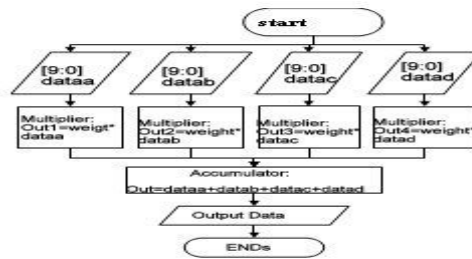Fig.5:.The flow of proposed neuron architecture

Fig.6:.Signal handling of multiplier accumulator

The architecture for the MAC is shown in Fig. 6. With tree configuration as shown in Fig.7, the use of tile logic is quite uneven and less efficient than with a chain. The idea of this configuration is that the 2 value from multiplier were added separately. The partition of the computation is then added at adder4 for the final output. Activation function in a back propagation network defines the way to obtain output of a neuron given the collective input from source synapses. The back propagation algorithm requires the activation function to be continuous and differentiable. It is desirable to have an activation function with its derivative easy to compute. The mathematical algorithm for tanh approximation using Taylor's Series expansion that is used in the hardware calculation is provided by equation 2: The design flow is presented in Fig 8.

$$y = x - x^3/3 + 2x^5/15 + \dots \qquad (2)$$
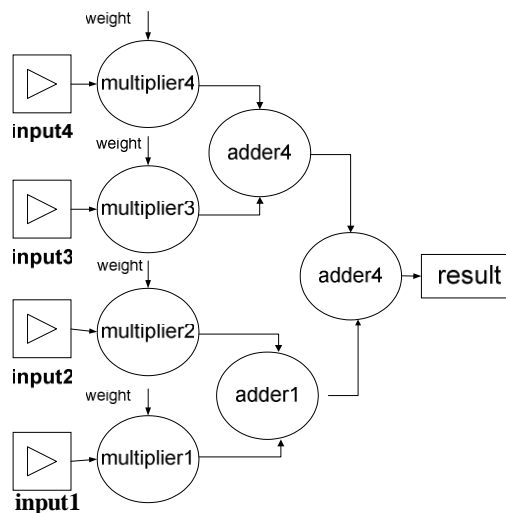
Fig.7:Tree configuration of Multiplier Accumulator
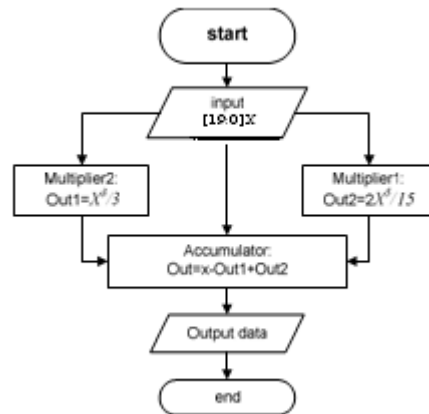
Fig.8:.Design flow of activation function

III. RESULTS AND DISCUSSION

3.1 : Simulation Results:

The following chapter consists of all the software and hardware results observed in the project. The results include snapshots of top module with the inputs, outputs and intermediate waveforms**.**
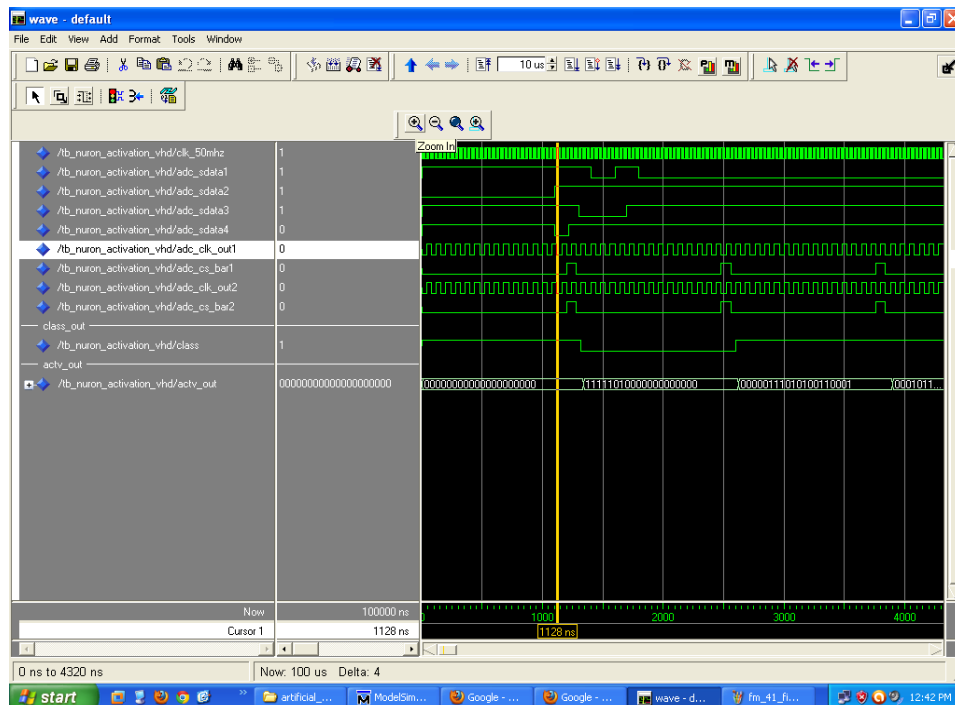

Fig.9.:Top Module simulation Results

3.2 . Chipscope Results:

Chipscope tool is used to view the Results after dumping the bit file into the FPGA. We need ICON(integrated Controller) and ILA(integrated logic analyzer) cores in order to run the chip scope tool. Chip Scope is an embedded, software based logic analyzer. By inserting an "integrated controller core" (icon) and an "integrated logic analyzer" (ila) into your design and connecting them properly, you can monitor any or all of the signals in your design. Chip Scope provides you with a convenient software based interface for controlling the "integrated logic analyzer," including setting the triggering options and viewing the waveforms.
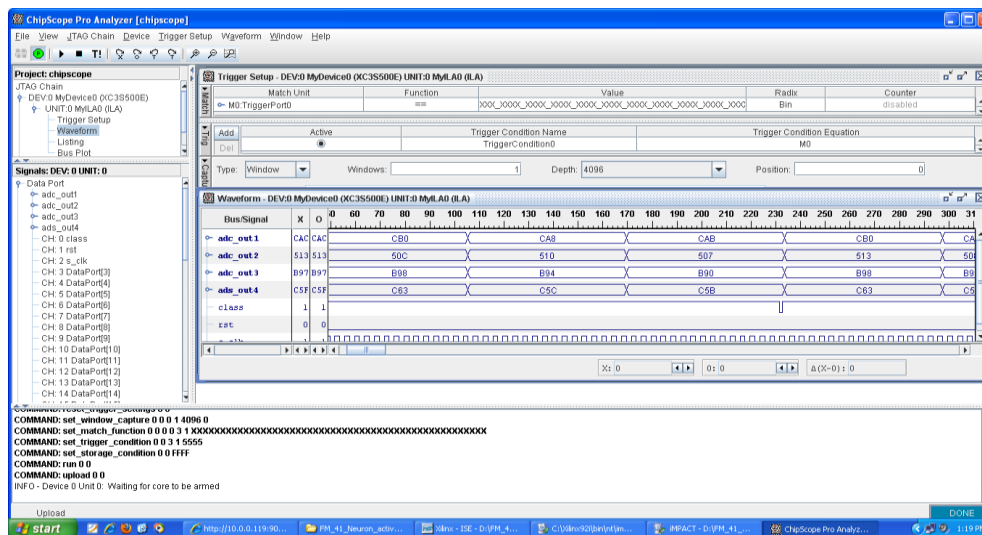
Fig.10.:Chipscope results

IV. CONCLUSION

The basic behaviour of the biological neuron can be emulated in an artificial neuron. A biological neuron with their dendrites, soma and axon can be characterized in an artificial neuron as a black box with inputs and an output. To implement the system the electronic pulses or spikes transmitted through neurons are replaced by digital signals or pulses. With all these things we get an electronic system that reproduces the behaviour of the biological neuron. The aim is to have the possibility of interconnect more of these artificial neurons to create a complete neuronal network. In this thesis work we only have focused our efforts to create an only artificial neuron. To complete the global system is not too complicate, to interconnect the artificial neuron with other neuron; the programmer should only connect the output of our neuron with the input of the next neuron, and in this way, with the other previous and next neurons. Besides, the programmer should codify a program that governs the relations between the different weights of the neuronal network, but this part is out of our analysis. With the VHDL code generated a FPGA can be programmed. Depending of the capacity of the FPGA used, a larger or less number of neuron can be programmed, depending on the same way of the number of interconnections of each neuron. Therefore, according to the results obtained, we can say that we have designed a system whose performance leads to the achievement of the objectives of the project and at the same time could work as the main base to develop future applications.

V. REFERENCES

[1]   Wan Fazlida Hanim Abdullah., Masuri Othman, Mohd Alaudin Mohd Ali,and Md. Shabiul Islam. Improving ion-sensitive field-effect transistor selectivity with back propagation neural   network. WSEAS Transactions of Circuits and Systems. 9(11): 700-712,2010

[2]   Stradiotto, N. R., Yamanaka, H., & Zanoni, M. V. B. Electrochemical sensors: a powerful tool in analytical chemistry. Journal of the Brazilian Chemical Society 14: 159-173, 2003

[3]   FPGA Implementation of a Multilayer Perceptron Neural Network using VHDL, Yamina TARIGHT, Michel HUBIN, Proceedings of ICSP '98 [page1-3]

[4]   'Overview and Use of the PICmicro Serial Peripheral Interface', Microchip(TM),[page 1-9]

[5]   Poole,D., Linear algebra : a modern introduction. Belmont, CA, Thomson Brooks/Cole, 2005

[6]   Polikar, R. Ensemble based systems in decision Making. IEEE Circuits and systems Magazine6(3)21-45,2006.

[7]   A.Durg,W.V.Stoelker,J.P.Cookson,S.E.Umbaugh and R.H.Moss,"Identification of Variegating Coloring in Skin Tumors:Neural Network vs Rule Based Induction methods",IEEE Eng. in med. and Biol., vol.12 pp.71-74 & 98,1993

[8]   FPGA Implementation of Artificial Neural Networks: An application on Medical   ExpertSystems,G.P.K Economou,E.P.Mariatos, N.M.Economopoulos, D.Lymberopoulos, and C.E. Goutis,Department of Electrical Engineering University of Patras, GR 261 10,Patras, Greece

[9]   A.R Ormondi and J.C Rajapakse, FPGA Implementation of Neural Network, [page 271-296],2006

[10]   Simon Haykin 'Neural Networks and Learning Machines, third edition, [page 40-45]

[11]   Benard Widrow, David E. Rumelhart, and Michael A. Lehr, "Neural networks: Applications   in   industry, business   and   science," Communications of the ACM,vol. 37, no. 3, pp. 93–105, Mar. 1994.

[12]   Altera Corporation 'Cyclone III Device Handbook, Volume, 1',[chapter5,page 1-8], July 2007

S. Sai Sree Andal[1] (M.Tech),

N. Aravind[2], Asst. Prof

Design and Fabrication of Mobile Phone Controlled Four Legged Walking Robot

Kadam Rohan Chandrakant[1], Mr. Vijayavithal Bongale[2], Mr. Sree Rajendra[3]

[1]*M.Tech. (IAR), Malnad College Of Engineering, Hassan (KA)*
[2]*Associate Professor, Dept. Of Mechanical Engg., Malnad College Of Engg., Hassan (KA)*
[3]*Associate Professor, Dept. Of Mechanical Engg., Malnad College Of Engg., Hassan (KA)*

ABSTRACT

In recent years, practical mobile robots have been successfully used in controlled environments such as factories, offices, and hospitals, as well as outdoors on prepared surfaces and terrain with minor irregularities. The most common type of robot has rigid body and is driven on wheels or tracks. However, for the operation of mobile robots in extremely rough, uneven terrain has been impossible or unreliable at best. So instead of wheeled robot, it has looked to the animal world for inspiration, attempting to develop walking robot to imitate the body structure and method of locomotion of mammals, human beings, and other arthropods. However, reliable mobility on extremely uneven terrain such as step climbing, gap crossing, gradients, side slopes remains an elusive goal for manmade devices. Most legged robots have been based on the concept of a single rigid body having articulated legs. Such a robot become attractive for any required application traversal of terrain that is difficult for wheeled and tracked robots, and is expensive or dangerous for humans.The present project work describes design of four legged robot which uses legs for its movement instead of wheels. This robot is capable of receiving few sets of command instructions in the form of DTMF tones and performs the necessary tasks. This robot is controlled by mobile and is capable of walking front & back and other specified direction with the help of the legs. The robot also senses the movement of living beings, leakage of gas and gives the siren and sends the message so that robot can be used as a security device in many applications.

KEYWORDS: *Animal world, Control system, DTMF, Legged robot, Sensor, Uneven terrain, Wheeled robot.*

I.   INTRODUCTION

Traditionally, most mobile robots have been equipped with wheels because the wheel is easy to control and direct and provides a stable base on which a robot can easy to build. One of the major drawbacks of the wheeled robot is that, it requires a relatively flat surface on which it can operate. On rocky or hilly terrain, which might be found in many applications as forestry, waste cleanup and planetary exploration, it is quiet difficult to operate with the use of wheels. A second approach to this problem would be to use tracked wheel robots. For many applications this is acceptable, especially in very controlled environments. However, in other instances the environment cannot be controlled or predicted and a robot must be able to adapt to its surroundings. Such a surrounding can be, places where robots would have to step over the obstacles such as a surfaces where pipes are running and where they have to move on the discontinuous terrain like steps. The research into legged robotics promises to overcome these difficulties. The complexity of control required for a legged robot to navigate autonomously over unfamiliar terrain has made them difficult to build. The use of legged robot is interesting for various applications, including cargo transport, entertainment, education, land mine removal, forestry and space exploration.

Their versatility allows the legged robot to access challenging terrains with high safety like paddy fields, sandy soils, and rough surfaces which are difficult for wheeled robot to cross. Similarly many manmade "terrains" are also difficult for robot because of narrow doorways, sharp turn, floor irregularities, ramps, steps, ladders, etc. So, in general, legged robot become attractive for any required application traversal of terrain that is too difficult for wheeled and tracked robots, and is too expensive or dangerous for humans. So this project work provides a legged robot which eliminates the limitations of wheeled robot that cannot operate on uneven terrain.

One can also think of detecting the human being by using the wireless remote controlled robot, which has the sensors that detects the presence of the human being and indicates the user. Also in hazardous situations the robot can be designed to detect the gas leakage using gas sensor. Using wireless robot, the unit can be easily mobilized and can be controlled by the micro controller, which is programmed to control the input and output modules interfaced to it. The controller makes use of a PIR and gas sensor to sense the human beings and any gas leakage respectively and give an alert indication through the buzzer. It also makes use of a mobile phone, which is used to control the robot.

The unit can have micro controller based motherboard which is present with the robot itself. It is interfaced with some servomotors for moving the robot, a PIR sensor for living organism presence detection, and a gas sensor to detect the leakage of gas, GSM module for sending the message and a DTMF decoder for receiving the instructions from the mobile phone utilizing eight servomotors. The motor generates torque directly from DC power supplied to the motor by using internal commutation, stationary permanent magnets, and rotating electrical magnets. The driver used for motors is L293D**.** The device is a monolithic integrated high voltage, high current four channel driver designed to accept standard DTL or TTL logic levels and drive inductive loads and switching power transistors. This project makes use of a micro controller, which is programmed, with the help of embedded C instructions. This microcontroller is capable of communicating with input and output modules. The controller is interfaced with DC motors, which are fixed to the robot to control the direction of the robot.

II.   HISTORY

Legged locomotion systems that have evolved in nature, show very good performances in terms of stability, payload capabilities, dynamic behavior. Thus, usually they are considered a very important source of inspiration for designing legged robotic systems mainly for aspects ranging from the mechatronic design to the path planning and gait generation. Several researchers have stressed these topics by using a multidisciplinary approach. For example, several studies have been addressed to the transmission system of vertebrate legged animals from a kinematic point of view. In fact, bones and articulations can be easily modeled as links and joints of a kinematic architecture. Those have been and still are an inspiration both for design and operation of walking legged systems. In different places of the world, there are many scientists worked on the different walking robot. E.g. RIMHO II,
SCOUT II, TITAN VIII, AIBO robot.

III.   METHODOLOGY
3.1 ARCHITECTURE OF THE PROJECT

Fig 3.1 explains the Architecture of the four legged robot
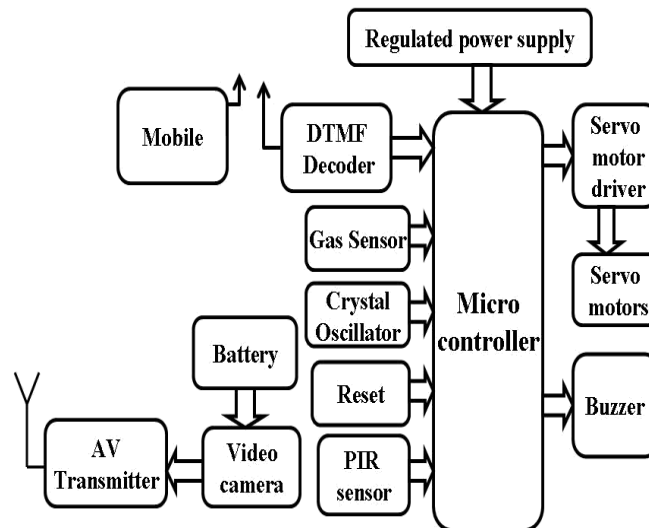

Fig. 3.1: block diagram of mobile phone controlled four legged walking robot.

3.1.1. Power supply: It consists of wireless communication; which requires 0 to 30 volt DC supply for the operation of the system.

3.1.2. Micro controller (16F877A): Microprocessors and microcontrollers are widely used in embedded system products. It is a programmable device. A microcontroller has a CPU in addition to a fixed amount of RAM, ROM, I/O ports and a timer embedded all on a single chip. The microcontroller used in this work is PIC16F877A

3.1.3. PIR sensor: Pyroelectric infrared (PIR) sensor is used for living organism detection. It provides an optimized circuit that will detect motion up to 6 meters away and can be used in burglar alarms and access control systems.

3.1.4. Gas Sensor: A MQ-6 gas sensor is used for combustible gas detection which with lower conductivity in cleans air. When the target combustible gas exists, the sensor's conductivity is higher along with the gas concentration rising.

3.1.5. DTMF Decoder: DTMF is the most common telecommunications signaling method used. DTMF stands for Dual Tone Multiple Frequency; is used to send information through phone lines to and from one local exchange. Most often, an MT 8870 or compatible circuit would be used as a DTMF decoder.

3.1.6. Servo motors and motor drives: There are eight servo motors used in this robot for the motion of the legs. For each leg two motors are required, one for knee and one for motion. Servo motor drives are used for the driving the motors.

3.2 Kinematic analysis:
The kinematic analysis of a four legged robot consisting of a fixed base and moving platform manipulated by four serial chain legs with six DOF per leg undergoing motion in space is presented. The serial chain legs in each configuration are kinematically identical to each other and are independently performed for each leg. Figure 3.2 shows the schematic of the four legged robot consisting of a fixed base and a moving platform. The rigid platform forms a plane in space and the orientation of every point of the platform is the same as that of the platform itself. The base of each of the four legs could be at some arbitrary distance from each other. The position and orientation of each leg relative to the base frame is then calculated.
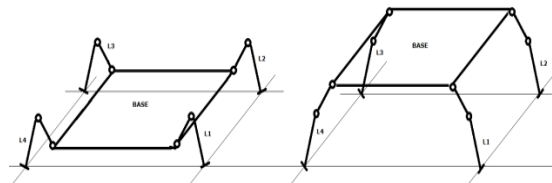
Fig. 3.2: Schematic of the four legged robot.
The robot kinematic diagram is given in Figure 3.3
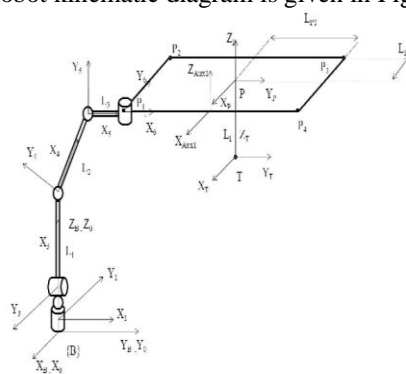
Fig. 3.3: Kinematic diagram.
The computation of joint angles θ1, θ2, θ3, completes the solution. Since these axes intersect and affect only the orientation of the manipulator, they can be computed based on the rotation portion as given below,

Eq. 20

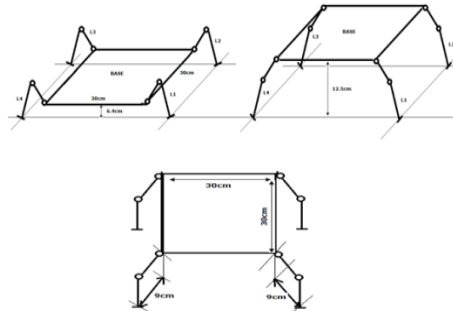$$\theta_1 = Atan2(-{}_3^0R_{13}, {}_3^0R_{23})$$

Eq. 21

$$\theta_2 = Atan2({}_3^0R_{33}, {}_3^0R_{23}/C_1)$$

Eq. 24

$$\theta_3 = Atan2(-{}_3^0R_{32}, {}_3^0R_{31})$$

3.3 Force and torque analysis:

Eq. 23

Fig. 3.4: Details of four leg mechanism.

III. SELECTION OF SERVO MOTORS

Weight of robot= 6 kg
Distance from servo output on foot (radius) = 9 cm = 0.09 m
Servo Torque = 0.6864 N-m
Weight=torque / radius= 7/9 = 0.78 ×9.8
   = 7.644 kg-f

Torque=	radius	×	weight
Torque=	0.09	×	7.644

Torque = 0.688 N-m
But that's only for whole body since there's at least 2 legs on the ground at a time so 0.688/2= *0.344 Nm* which means that standing still, the robot should be able to easily support the weight. So *static*= 0.688N-m, then no need to Figure out dynamic. But for that it needs to get the acceleration of the robot so depending how fast the robot to accelerate, my motors may require more torque.
In moving condition the robot covered the 30 cm distance in 4 sec. so, distance covered in 1sec = 7.5 cm, therefore the acceleration of robot is 7.5 cm/s² i.e. 0.075 m/s²

Force=	mass	×	acceleration

Force= 6 kg × 0.075 m/s²
Force= 0.45 N
Dynamic torque = force × radius
                 = 0.45 × 0.1
                 = 0.045 N-m
Then:

Torque	=	static	+	dynamic

Torque = 0.688 + 0.046
Torque = 0.752N-m
Then:

torque	=	0.752/2

torque = 0.376 N-m   (for each leg)

IV. ADVANTAGES

➢   They are capable of omnidirectional motion without the need for turn-in-place.
➢   They can access a wide variety of terrains. Boulder fields, steep slopes and loose, sandy areas can all be traversed by walking, in addition to any location accessible to wheeled robots.
➢   A legged system is well adaptive to uneven terrains, namely the legs can be arranged according to the level changes.
➢   The legged robots can be used to reach in both structured and unstructured environments
➢   Legged robots can negotiate irregular terrain while keeping their body always levelled. This is important when carrying on board sensors and pieces of equipment that need to be levelled.

V. CONCLUSION

It is designed such that the robot can be operated using mobile phone in any part of the world. It is designed with PIR sensor, Gas sensor and a video camera mounted on it so it can be used,

➢ To detect persons in restricted areas.
➢ To detect the human beings and detect the gas leakage in the hazardous condition of mine (i.e. Coal mine, gold mine, etc.).
➢ In industries to detect the several gas leakages and human detection to avoid the human casualties in any accidental condition where a man cannot be reached.
➢ In forest to detect living animals without any person going to the forest and search for the living animals.
➢ At home as safety device to detect the gas leakage.
➢ In the army as spy robot.

REFERENCES

[1]. Gonzalez de Santos, Pablo; Garcia, Elena; Estremera, and Joaquin. *Quadruped Locomotion: An Introduction to the Control of Four-legged robots*. Springer-Verlag New York Inc, 7 2006.
[2]. S. Peng, C. Lani, and G. Cole, "A biologically inspired four legged walking robot," in International Conference on Robotics and Automation, 2003, pp. 2024–2030.
[3]. H. Kimura and Y. Fukuoka, "Adaptive dynamic walking of the quadruped on irregular terrain - autonomous adaptation using the neural system model," in International Conference on Robotics and Automation, 2000, pp. 436–443
[4]. Harris S.E., "Horse Gaits, Balance and Movement", Howell Reference Books, New York, 1993.
[5]. C.-L. Shih and C. A. Klein, "An adaptive gait for legged walking machines over rough terrain," IEEE Trans. Syst. Man Cybem., vol SMC-23, no.4, pp. 1150-1 155, July/Aug. 1993.
[6]. R. B. McGhee and G. I. Iswandhi, "Adaptive locomotion of a multi legged robot over rough terrain,"IEEE Trans. Syst. Man Cybem. Vol. SMC-9, no.4, pp. 176-182, Apr. 1979.
[7]. Hirose, Shigeo; Kato, Keisuke. (1998).; "Development of Quadruped Walking Robot with the Mission of Mine Detection and Removal – Proposal of Shape-Feedback Master-Slave Arm –"In: *Proceedings of the 1998 IEEE International Conference on Robotics and Automation*; pp 1713 – 1718.

Fem Based Analysis Of Chip Tool Interactions To Study The Stress Distribution On The Rake Face

[1]Mr.G.Balamurali, [2]Mr. Bade Venkata Suresh,
[3]Mrs.Y.Shireesha, [4]Mrs.T.Venkata Sylaja
[1,2,3,4]*PG Student, Dept of Mechanical Engineering GMR Institute of Technology*
Rajam, Srikakulam, A.P, India

ABSTRACT

Introduction of green concepts in machining operations is being envisaged by introducing different echo friendly cooling systems in the modern machine shops. The role of cutting fluids usage in metal cutting is predominant as it influences the surface quality and production cost. The current work mainly focuses on the study of chip tool interactions viz. contact pressure, temperature and chip flow pattern on the rake surface in plain turning operation for different cutting parameters without any cooling medium and analyze the influence of high pressure air jet as the cooling medium on the chip tool interactions like contact pressure reducing the tool wear, cutting temperatures thereby increasing tool life.

KEYWORDS: *Modelling; Machining; Compressor; Air Jets; Nozzle;*

I. INTRODUCTION

The use of high speed air jet as a coolant in machining is a challenging scenario in environmental friendly machining. Despite the extensive literature, air jet cooling in machining is an area of ongoing research. Until now, the jet cooling technique has been studied only from a thermal point of view. The new aspect investigated in this work is the chip bending ability of the jet. The idea of chip-bending and its beneficial effects in cooling the cutting area is not related to maximizing the heat transfer, but to avoid the temperature increase. The heat generation in the chip-tool interface is due to the contribution of deformation in the shear zone and to the frictional contact between the chip and the rake face of the cutting tool. The importance of the frictional contact is proportional to the friction coefficient and to the pressure of the chip on the rake face. The traditional way of reducing this contribution is using a cutting fluid (flooding) or, more recently, injecting a coolant in the chip-tool interface. The new approach with high speed air jet shows the temperature reduction is strongly dependant on the position of the nozzle. By directing the jet onto the top face of the chip it is possible to reduce the pressure on the rake face, responsible of temperature increase in the chip-tool interface. The pressure on the top face of the chip generates a stress on the bottom face of the chip close to the constraint and in the chip-tool interface. The global stress is due to air jet pressure and cutting pressure on the rake face. When the air jet is directed on the top face of the chip (overhead position) the global stress is less than the cutting stress in dry machining. A fully thermo-mechanical model has been developed with DEFORM-3D and a mechanical only model with DEFORM-2D, in order to investigate the chip bending. From an analytical point of view the chip can be modeled as a structural cantilevered beam with uniform load. The results from finite element modeling show the displacement of the chip is mainly due to the chip-breaker. The displacement due to the air jet bending moment is minimized by the stiffness close to the constraint point, but the mechanical effect of the air jet has a significant impact on the energy in the tool.

II. DRY CUTTING MODE

A. Objective

Analysis of effect of the cutting parameters like cutting speed, feed rate and depth of cut on cutting force components which influence the contact pressure, temperature and chip flow pattern on the rake surface during turning operation

B. Equipment

Lathe machine, Lathe Tool Dynamometer, Amplifier, Cutting tool, PC, Job piece.

C. Experimental Setup

Figure 1 shows the schematic of the experimental setup for carrying out the experiment. Work piece is mounted in the chuck of the lathe headstock. The tool dynamometer is mounted on the carriage at the place of tool holder[6].The tool holder is mounted on the dynamometer as shown in Figure 2. Output of the dynamometer is amplified by charge amplifier (Kistler 5070A) and data are collected in the PC by using data acquisition system[7]. This setup is used to find the cutting forces for different speeds feeds and depth of cuts. The results are carefully tabulated and are used for the analysis of the cutting tool ,stress distribution on the rake face of the tool by the forces that are taken from the dynamometer
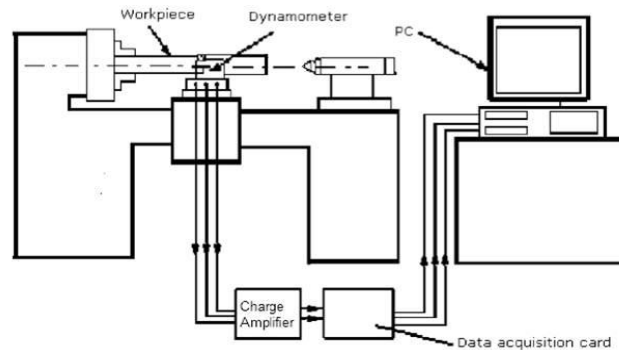
Fig 1:   Schematic diagram of the experimental setup.
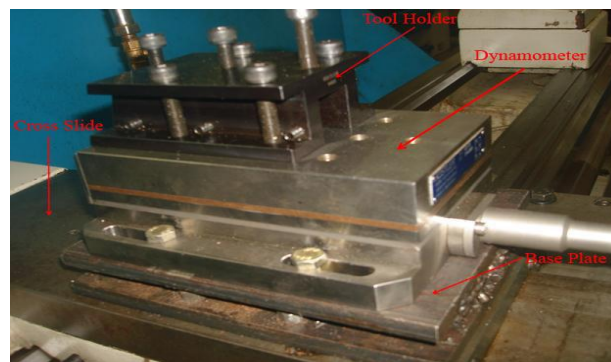
Fig 2:   Eexperimental setup for finding the cutting forces

D. Observations of Cutting forces

Initial diameter of the bar = 25mm
Bar material = MS(Mild steel)
Cutting tool material = HSS(High speed steel)[2]

TABLE 1:   OBSERVATIONS OF CUTTING FORCES WITH OUT COOLANT

Speed (rpm)	Feed (mm/rev.)	Depth of Cut (mm)	Cutting Force (N)
550	0.1	0.1	23.88
440	0.1	0.1	17.73835
330	0.1	0.1	13.9618
220	0.1	0.1	22.2015
118	0.1	0.1	14.57215
118	0.2	0.1	20.82825
220	0.2	0.1	21.293
330	0.2	0.1	19.0354
440	0.2	0.1	30.9735
550	0.2	0.1	16.86095
550	0.3	0.1	67.5965

440	0.3	0.1	44.5175
330	0.3	0.1	36.3159
220	0.3	0.1	23.68925
118	0.3	0.1	62.866
118	0.4	0.1	46.0434
220	0.4	0.1	44.632
330	0.4	0.1	24.719
440	0.4	0.1	28.8391

III.   MATERIAL PROPERTIES

_____

Material	:	High Speed Steel
Young's modulus	:	190-210Gpa
Poisson's ratio	:	0.27
Density	:	7800 kg/m3
Work piece	:	Mild steel

_____
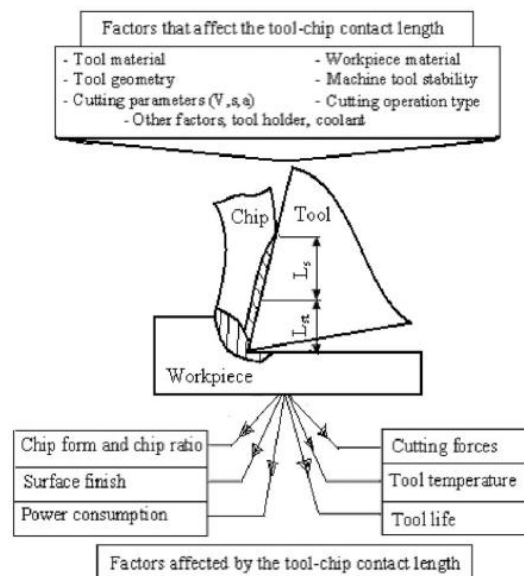
Fig 3:   Chip Tool contact length

IV.   ESTIMATION OF THE CHIP TOOL CONTACT LENGTH

A number of theoretical and experimental estimators have been proposed for the contact length in the orthogonal cutting process Based on the experiments conducted on different types of steel using a tool with an unrestricted rake face, a relationship between the chip–tool contact length, chip thickness, the chip compression ratio and the friction coefficient has been developed. It suggests that the length of the sticking region is approximately equal to the deformed chip thickness hc, and in accordance with Tay's assumption[3]
Total chip–tool contact length Lc as shown in the Figure 3 is given as [4]

_____

Lc	=	2hc
Thickness of the chip	=	1.5mm
Width of the chip	=	4mm
Contact Length	=	2*1.5 =3mm

Area of the chip contacting the tool =3*4= 12mm^2
Pressure acting on the contact area =Force/Area
P1   =23.88/12   =1.99 N/mm^2
Similarly for the remaining forces
P2   =1.419 N/mm^2
P3   =1.163 N/mm^2
P4   =1.85012 N/mm^2
P5   =1.214 N/mm^2

_____

V. ANLAYSIS OF CUTTING TOOL WITH OUT COOLENT

A. *For the Speed (N) =550 rpm, Feed=0.1mm/rev, Depth of cut=0.1mm*

1) *Meshing*

Finite element method is purely based on the dividing the cutting tool in to finite number of elements .Here the edge length that is taken for division of elements is 5mm.The mesh division is as shown in the Figure 4
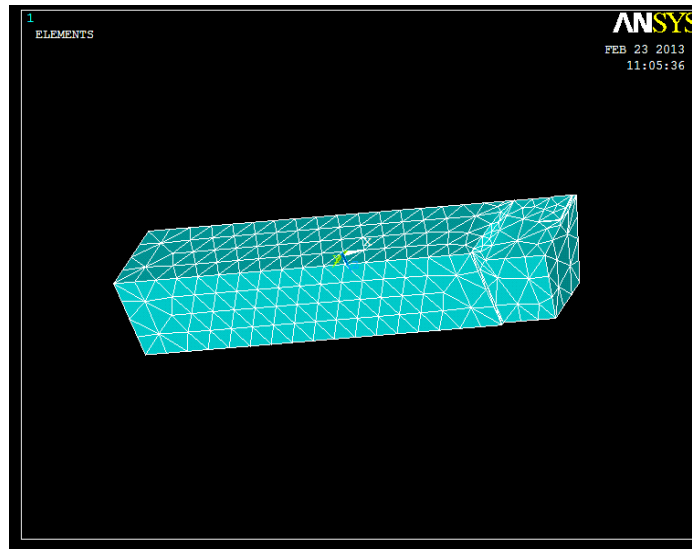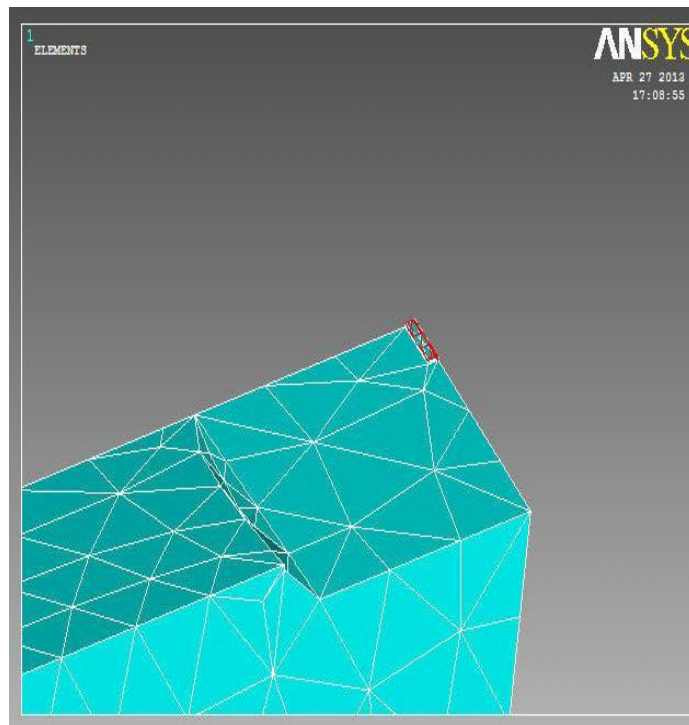

**Fig 4 Meshing of cutting tool**

2) **Area Selection For Applying The Load**

For applying the load that is pressure first choose the area that the chip is in contact with the tool .Then apply the load on that area as in the Figure 5

Fig 5: Area selection for applying the load of cutting tool

3) *VonMises Stress*

After load is applied by giving the material properties and constraints on the selected area, deformation and stresses are developed .The required Von misses stresses are show in the Figure 6
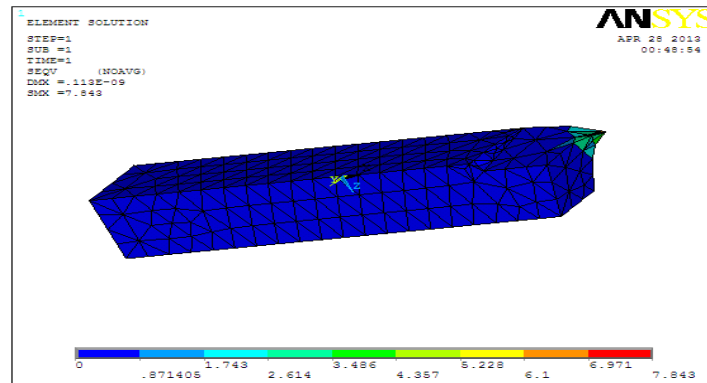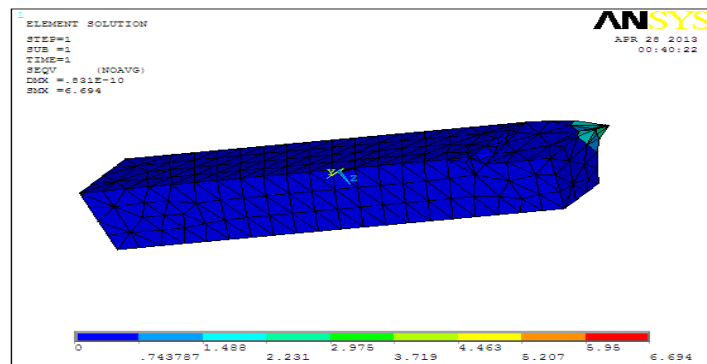

Fig6 VonMisses Stresses of cutting tool

Maximum deflection is 0.151e^{-09} mm.
Maximum Stress is 3.21 N/mm^2.
Now for another different set of speed, feed and depth of cut Von misses stresses are taken by using ansys and are shown in the below Figure 7.
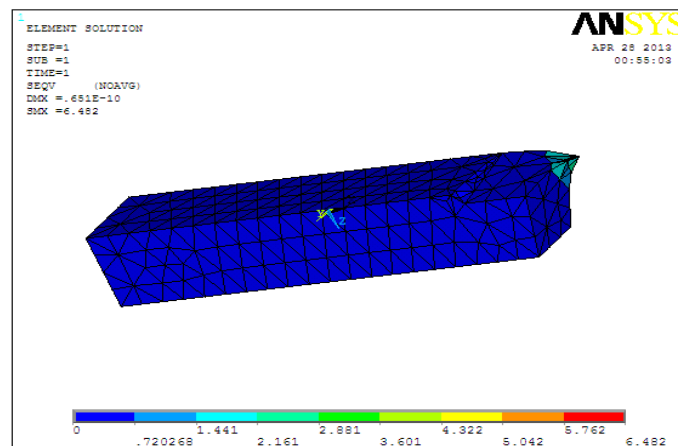**B.** *For the Speed (N) =440 rpm, Feed=0.1mm/rev, Depth of cut=0.1mm ,pressure=1.419 N/mm^2*

Fig 7: Von misses stresses for cutting tool

Maximum Stress 2.289N/mm^2
C. *For the Speed (N) =440 rpm, Feed=0.1mm/rev, Depth of cut=0.1mm*

Fig 8: Von misses stresses for cutting tool

Maximum stress is 1.876N/mm^2

A graph is drawn between the pressures Vs Stress. The stresses which are taken from the applied pressures on the cutting tool, Stress varies linearly with the pressure as shown in the Figure 9
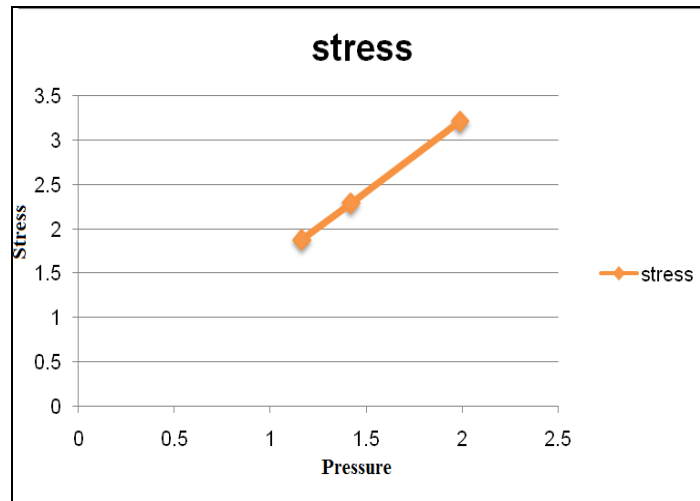

Fig 9:   Pressure Vs. Stress Graph

D.   TOOL LIFE

Taylor Tool Life Equation [1]

$VT^n=C$

Where   v = cutting speed, m/min;

T = tool life, min;

n and C are parameters that depend on feed, depth of cut ,work material, and tooling material but mostly on material (work and tool).

TABLE 2:   TYPICAL VALUES OF N AND C

Tool material	n	C
High Speed Steel		
Non Steel Work	0.125	120
Steel Work	0.125	70
Cemented carbide		
Non Steel Work	0.25	900
Steel Work	0.25	500
Ceramic		
Steel Work	0.6	3000

E.   TOOL LIFE CALCULATIONS[5]

Cutting velocity=πDN/1000 (m/min)

D=Spindle diameter (mm)

1)   Sample calculation

V1=πx25x550/1000

= 43.19 m/min

Similarly

V2= 34.55 m/min

V3=25.19 m/min

V4=17.29 m/min

2)   Tool Life

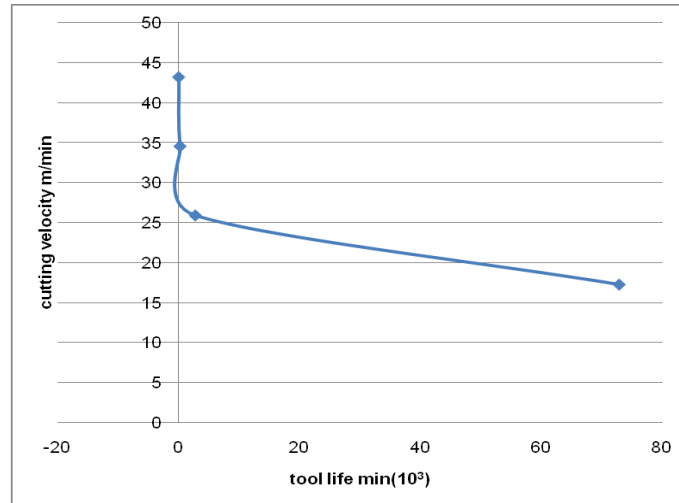$VT^n=C$

T1=   (70/43.19)(1/0.125)

=   49.32 min

Similarly

T2=   283.92 min
T3=   2838.2 min
T4=   72.85x103 min

Fig 9: Graph between Cutting speed Vs Tool Life

A graph is drawn between the cutting speed Vs Tool life which varies parabolic shape .As the cutting speed increases tool life decreases as cutting speed decreases tool life increases and tool life slightly decreases at low cutting speeds which is as shown in the Figure 9

CONCLUSION

Modeling and analysis of cutting tool which is used in the turning operation without cutting fluid ie in dry cutting mode is done by the collecting the data of cutting forces that are acting on the rake face by the use dynamometer .By the modeling and analysis stresses are determined ,to study the wear pattern and a system is developed to reduce the tool wear

ACKNOWLEDGMENT

REFERENCES

[1].   Leone, W.C.; "Distribution of shear zone heat in metal cutting"; Transactions of ASME, vol. 76, pp. 121-125, 1954
[2].   Adibi-Sedeh, A.H., Madhavan, V., Bahr, B.; "Extension of Oxley's analysis of machining to use different material models"; ASME Journal of Manufacturing Science and Engineering, vol.125, pp.656-666, 2003
[3].   Leowen, E.G., Shaw, M.C.; "On the analysis of cutting tool temperatures"; Transactions of ASME, vol.71, pp.217-231, 1954
[4].   Jaeger, J.C.; "Moving sources of heat and the temperatures of sliding contacts"; Proceedings of Royal Society N.S.W., vol.76, pp.203-204, 1942
[5].   Smithey, D.W., Kapoor, S.G., DeVor, R.E.; "A new mechanistic model
[6].   For predicting worn tool cutting forces"; Machining science and
[7].   Technology, vol.5/1, pp.23-42, 2001
[8].   Shaw, M.C.; "Metal Cutting Principles"; 1984; ISBN 0-19-859020-2

SAFETY ENGINEERING

[1]SANGOTOLA T.M., [2]FOLAMI, F.T. AND [2]OPALEYE, E.T.

1department Of Civil Engineering, 2department Of Electrical Engineering
The Polytechnic, Ibadan.

ABSTRACT

Safety is the state or condition of freedom from danger or risk. It cannot be restricted to a field alone it is what's applicable to us in our day to day activities. This work covers safety in large organizations in relation to personal, fire and gas and automobile. Other areas which this work cover are safety at the place of work, safety of site in relation to protection of an individuals from the consequences of its own actions, safety of road men from potential hazard on a particular site (electricity, examinations and falls) and safety of road men from another's actions, regulations and rules as it relates personal, safety habit, workshop safety habits, tools safety, workshop machines tools safety and electrical equipment safety. This work sheds light on the characteristics of fire and the causes of fire outbreaks in public buildings (office buildings, hotels, cinema halls, warehouses etc). Also, it gives details on fire protections, precautions and means of escape in case of fire.

KEYWORDS*: Safety, Environment, Equipment*

I.   INTRODUCTION

From the cradle to the great people are exposed to a great variety of hazard based on the environment in which they carry out their daily activities. At the recent time due to technological development, man's work becomes more sophisticated than before in terms of materials, equipment and machineries. There are hazard in the street, at home and in factories/workshop work in the utility field is potentially more hazardous than in other occupation but confirming to safe practices, one can attain working in a more manageable harmless working environment. Safety is an act of incubating the necessity of taken precaution for the avoidance or reduction of accidents in order to protect people and property. Safety in work place is concerned with all the safety mechanism put in place by employer individual to ensure, as much as possible, the avoidance or elimination of accident in industries or work place. Okorie (2000).Generally, safety is required in any daily activities involving human being either at work or at home. It is equally important that all reasonable worker or human being at work should always place, priorities or safety before performing any functions or duties which could be terms as "SAFETY FIRST". This idea is applicable in some organizations but still need to be practice by each individual in any environment.

II.   SAFETY AT THE PLACE OF WORK

Both the employers and employee had to cooperate to enable a safe working place to exist. In a construction site supervisor will need to;
[1]   Identify potential hazards that could exist on the site and take the necessary precautions.
[2]   Be vigilant in ensuring that young under 18 years and holder men are not placed upon work that is new to them they must be adequately trained and supervised on their work.
[3]   Ensure that all practices on site confirm with appropriate legislation (law).
[4]   Attempt to improve on the minimum – requirement by say, encouraging a dean, orderly is – well stake material and clear access route.

SAFETY OF SITE

The following safety should be considered when taken precautions to ensure safety on a highway site.
[1]   The production of an individual from the consequences of his own action.
[2]   The safety of road men from potential hazard on a particular site.
[3]   The safety of road men from one another's actions.
[4]   The safety of the traveling public.
[5]   The protection of road men from the traveling public.

THE PROTECTION OF AN INDIVIDUAL FROM THE CONSEQUENCES OF ITS OWN ACTION

Ignorance of the work process involved or carelessness will cause a man to injure himself through his own actions. Many accidents could be prevented with the use of protective equipment such as;

a. Protective spectacles, goggles or visor when working with hammers and chisels or cutting machineries at concrete or stone.
b. Helmet when working below other work men or in situations where object are liable to fall from above.
c. Protective boots with steel caps and soles will prevent injuries through crushing and treading on nails.
d. Gloves in manual activities such as curbing etc.
e. Face mask when cutting concrete or stone with disk cutters.
f. High visibility and reflective jackets for men working on the highway.
g. Far muffs are plugs when using compared air tools such as breakers and drills and also with pilling equipment.

III. THE SAFETY OF ROAD MEN FROM POTENTIAL HAZARD ON A PARTICUALR SITE ELECTRICITY ON SITE

[1]   **Underground cables:** Some cables are armored by a wrapping of steels tape but many have only light covering which can easily be damaged by excavators or with hard tools like pickaxe when buried at depth between 450mm to 1m. First check the electricity had been disconnected from mains when demolishing building.

[2]   **Services in vicinity of site:** Before and during precaution must be taken to prevent accidental contact with live underground or overload cable. Disconnecting or divert the electricity supply authority's cable.

[3]   **Supply for site:** When temporary lightning is installed or portable tools are used the risk of serious injuries can be reduces by wings a transformer to reduce normal supply voltage of 250v to 230v or less.

[4]   **Overhead electricity lines:** They are usually insulated and if a metallic object is brought into contact or in close. Proximity with an overhead conductor then an electric current will discharge through the object to earth, with this precautions should be taken to prevent crane jibs,scaffold poles or vehicle tipper bodies getting too dose. A method is to erect "good posts" which will channel plant along a particular route and through the posts.

EXCAVATIONS

Many dangers arise from excavation. The principle danger is that ofthe collapse of the site of an excavation. Since all ground will collapse under certain conditions an excavation should be clearly supported (timbering) or the sites should be sloped to a safe and angle as soon as the excavation reaches a depth where men could be buried or trapped should a collapsetake place.

FALLS

Almost half the accidents in construction work are done to the falls of men or materials. Men not only fall from working places, they slip and fall when making their way from one part of the site of the part to another and can also be injured when badly stacked materials fall on to them.
To prevent persons from falling:

1.   Provide hand rails for stairs and gangways over trenches.
2.   Check that any safeguard, such as guardrails, tailboards, handrails andcovers over opening in floors, which have been provided are kept in place.
3.   Keep the site tidy and get rid of rubbish, stove materials methodicallyand establish door routes through the site.
4.   Fence or cover all openings in floors, holes and ground and similarhazards.
5.   Provide good lighting on stairs, passages and other access route.
6.   Use only well - constructed ladders properly secures, suitable scaffolding for all work that cannot be done safely from ground.

To prevent materials from falling:
1.   Material should not be through down haphazardly from a height lower than or provide a chute or if material are to be thrown down off the area into which they will fall.
2.   Ensure that materials are properly staked and that they are not likely to be blown or knocked over.

3. If persons are working regularly in places where they are liable to be stuck by falling materials, provide a strong protective cover. Similar protection is advisable over passageways in common use. Provide safety helmets.

IV. THE SAFETY OF A ROADMEN FROM ANOTHER'S ACTIONS.

Many dangers arise from the thoughtless actions of men unconcernedwith the safety of their workmates. Some of the move common and those resulting in high incidents are listed below:

a. Trading on nails projecting from lose timber in a common site accident.
b. Not reporting defects in machinery and equipment resultingin unsafe plant being used by others.
c. Creating makeshift and unsafe situation such as sub-standard electrical connections and trailing cables etc.
d. Failure to replace guarding equipment after carrying out workwhich necessitates it removal.
e. Careless operation of machinery without ensuring proper safeguards such ashelp in reversing or use of a banks man with cranes or earthmoving machinery.

SAFETY RELATION TO PERSONNEL IN AN ORGANIZATION

In large organization induction courses are organized for new staff in order to carry out his duties which are administered by the personnel department. This induction will cover policies of the organization, chain or command, introduction to company mammals, introduction to equipment to be used. Sometimes at the induction the new employee will be acquired.

SAFETY IN RELATION TO FIRE AND GAS

Firefighting (fire extinguisher) is recommended in all laboratories workshops, buildings vehicles, rigs etc. in large organization personnel are trained on how to use fire-fighting equipment during fire drills and training exercises. The fire extinguishers shall be marked with an identification symboland the letter(s) that indicate the class of fires they shall be used for.

The fires shall be treated depending on their class:
CLASS A FIRE: Occurs in common materials such as wood, paper mattresses, rag, rubbish etc. the quenching and cooling effect of water or water solution is recommended.

CLASS B FIRES:occurs in the vapour-air mixture over the surface of flammable liquid such as gasoline, oil, grease, paints and thinners. Class B fires are extinguished by limiting air (oxygen) or by providing combustion-inhibiting agent.

CLASS C FIRE: Occur in or near electrical equipment non-conductiveextinguishing agents must be used.

CLASS D FIRES: Occurs in combustive metals such as magnesium,titanium, zirconium, lithium and sodium. Specialized techniques, extinguishing agent and equipment must be to control this type of fire. Generally all vehicles supposedto have a road-book, indicating the mileage and the maintenance performed. To avoid unnecessary accident the driver is expected to conduct a daily inspection of his/her vehicle, what should be checked are:

1. Engine oil level
2. Engine cooling fluid level (water in radiator)
3. Tyres condition and pressure
4. Automotive lights (full, dim, brake, parking, reverse)
5. Spare tyre and jack
6. Automotive documents
7. Windscreen document.

In vehicle seat belts are supposed to be worn by all passengers prior starting the engine (for front rear seats).Pregnant women who might feel discomfort from the seat belts and children under 10years old shall not be allowed in the front seat. It is also very important that the driver must have driving license recognized by the local authority. Also drivers are expected to be trained in defensive driving. The driver should not be under influence of alcohol (more than 0.3 gll), narcotic or drug is forbidden. The driver is supposed to knowthe driving speed for some areas and roads for example:
Driving speed shall not exceed in any case:

- Paved roads outside build up areas………………………. 120kph
- Gravel roads …………………………………………….   70kph

- Build up area ………………………………………….. 50kph
- Work site ………………………………………………… 5kph

Driving at night should be avoided as much as possible because of Visibility.

PERSONAL SAFETY HABITS

1. Remove loose fitting outer garments
2. Keep long hair of eyes
3. Remove or tuck tie in, close to collars
4. Roll up long sleeve shirts above elbows
5. Wearwrist watches and jewelleries when using machine
6. Wear shoes with strong toe cap
7. Wearprotective clothing e.g. close fitting apron tied at the back

WORKSHOP SAFETY HABITS

1. Walkand do not run in the workshop
2. Make sure your working space is clear of pieces of wood and stone
3. Always wash your hands before and after work
4. Note the positions of fire aid box and fire extinguishers.
5. Keep tool and equipment in appropriate places immediately after use
6. Be sensible to fooling around and no noise making in the workshop.
7. Report ay injury, no matter how minor it may be to the personnel in theworkshop
8. Remove paint, oil and grease from the floor immediately.

HAND TOOLS SAFETY

1. When passing sharp edge or pointed tools to another person pass them through by the handles.
2. Do not carry too much tools at once.
3. Store tools kits or in correct places immediately after use
4. Do not sharp edged tools of pointed tools without protecting the ends.

WORKSHOP MACHINE TOOL SAFETY

1. Observers should stand at a safe distance away from the machine.
2. Observe correct workshop clothing rule and wear ear plugs and goggles or eye shield.
3. Make sure there is adequate working space to operate the machine safety.
4. Do not interfere with switches or controls of a machine without permission.

ELECTRICAL EQUIPMENT SAFETY

1. Check the plugs frequently for fracture
2. Do not use damp or faulty lead
3. Do not use crack or faulty plugs
4. Store power tools and lead in a dry place when not in use
5. Report any damaged leads, plugs or switches to the right personnel
6. Always making sure your hands are dry and don't stand on damp floor when plugs, of equipment after use.
7. Switch off and disconnect plugs of equipment after use
8. Do not drag electrical cable on the floor

CHARACTERISTICS OF FIRE

Fire or combustion is the process of burning, it is a chemical reaction in which a substance combines with oxygen in the air and the process is accompanied by the emission of heat light and sound. Three things are essential before a fire can occur; they are a combustible substance, oxygen and initial source of heat. Remove anyone of these and the process of fire will cease.

Fire extinction consists the limitation of one or more of these factors as explained below.

a. Starvation (or the limitation of fuel)
b. Smothering (or the limitation of oxygen)
c. Cooling (or the limitation of temperature)

Causes of fire incidents in public buildings.

Human factors which bother on moral and physical hazards have been identified as causes of fire incident in public buildings. However, a good number of reported cases of fire outbreak in public building (especially in Nigeria) are such that the main causes of the fire are not made public. This is a discredit as it due

not allow the people to learn from the experience of the past so as to know and learn how to avoid the causes of fire outbreak.

The under listed are the major causes of fire in various public building:
a. Careless handling of fire e.g. careless smoking of cigarette
b. Enemy actor or arson
c. Rubbish burning
d. Electric/mechanical spark
e. Improper maintenance of machineries or engines
f. Lighting, tremors and earthquake - natural causes
g. Bad Storage offlammable liquids and other combustible material
h. Chimneys - Any defect in the basic principle of operation of it.
i. Spontaneous Combustion - e.g. reaction of quicklime with water in laboratory.
j. Naked light e.g. welding and cutting operation or use of blow lamps.
k. Friction - e.g. tension adjustment on belt driven machinery, serious over heating can be caused if the belt is to loose or too tight.

FIRE PRECAUTIONS

Fire can be prevented by using the following means:
1. Smoke detectors - this gives a quicker response time than heat defector as fire outbreak will often give a pulse of smoke or combustion products in early stage of five developments before rise in temperature is apparent.
2. Automatic fire extinction - having considered automatic detection it is logical to move into automatic extinction.
3. Automatic water systems: it provides a convenient form of protection which can operate without human factor. Water may be applied in form of coarse droplets or as a mist of fog and supplied by a system of mains and branch pipeswhichmaybeentheirwetordry.

V. RECOMMENDATIONS

It is advisable that a man must be trained or make familiar with any techniques before being allowed to begin work.When timber framing or shuttering or dismantled, nails should be removed or knocked flat or the timber should be stacked where it cannot be trodden on. Almost half the accident in construction works are due to thefalls of men or materials, men not only fall from working places, they slip and fall when making their way from one part of the site to another and can also be injured when badly stacked materials fall to them.To prevent person from falling keep the site tidy and rid of rubbish store material methodologically and establish clear routes through the site.Provide hand rails for stairs and gangways over trenches. Fence orcomer all openings in floors, holes in grand and similar hazards.Provide good lighting on stair, passages and other access routes use only well construction ladder properly secured, and where necessary, check that any safeguards, such as guardrails, toe boards covers over openings infloors which have been provided are kept in place for properly safety precautions.For any reportable accident, an accident investigation shall be carried out on site and an analysis shall be made to determine the action(s) to be taken for preventing re-occurrence.

VI. CONCLUSION

Far too often, rather than actually influencing the design safety engineers can be assigned to prove that an existing, completed design is safety problems late in the design process, correcting them can be very expensive. This type of error has the potential to waste large sums of money so it should be avoided.There is great need to improve fire services allover the state of the federation. Thousands of lives and properties would have been save if the department have been functioning well.The issues of transportation and equipment in case of fire outbreak can be solve by providing modem vehicle and modem firefighting equipment to replace obsolete one still in use now.

REFERENCES

[1]. Arthur Wiguall, Peters. Kendrick (1982):Roadwork Theory and Practice (1st Edition) London: Publishing in Association with the Institution of Works and Highways Technician Engineers by Heinemann: London.
[2]. Aremu J.A, Olaoye I.O., Ariwoola I.A., Asimi M.A, (2001): Introduction to Wood Technology (1st Edition).
[3]. Feyon A., Bobillier A.: Health, Safety and Environment Manual (3rd Edition).
[4]. Okorie J.U (2000): Developing Nigeria's Work Force (1st Edition)
[5]. Walton, John A., (1979): Wood in Theory and Practice (5th Edition).
[6]. Olayinka A.R and Ojo O.S (1996): Fire Safety in Public Building Company (1st Edition).