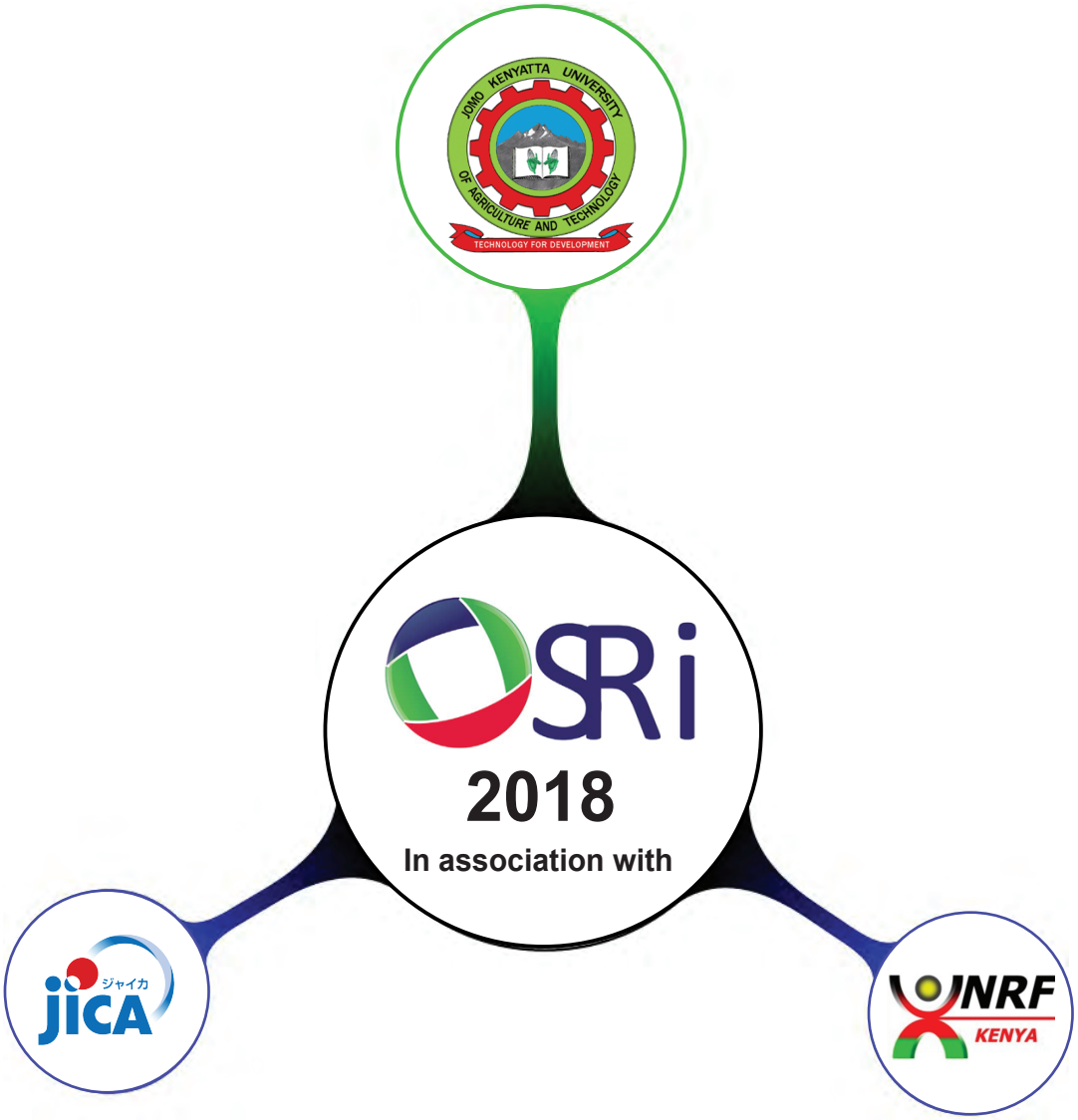


Proceedings of the 2018 Annual Sustainable Research and Innovation (SRI) Conference

2 - 4 May, 2018

JKUAT Main Campus

Juja, Kenya

ISSN: 2079-6226

Proceedings of the
2018 Annual Sustainable Research and Innovation Conference
JKUAT Main Campus
Juja, Kenya
2 - 4 May, 2018

The final version of the proceedings will be made available on;
http://sri.jkuat.ac.ke/ojs/index.php/proceedings/index

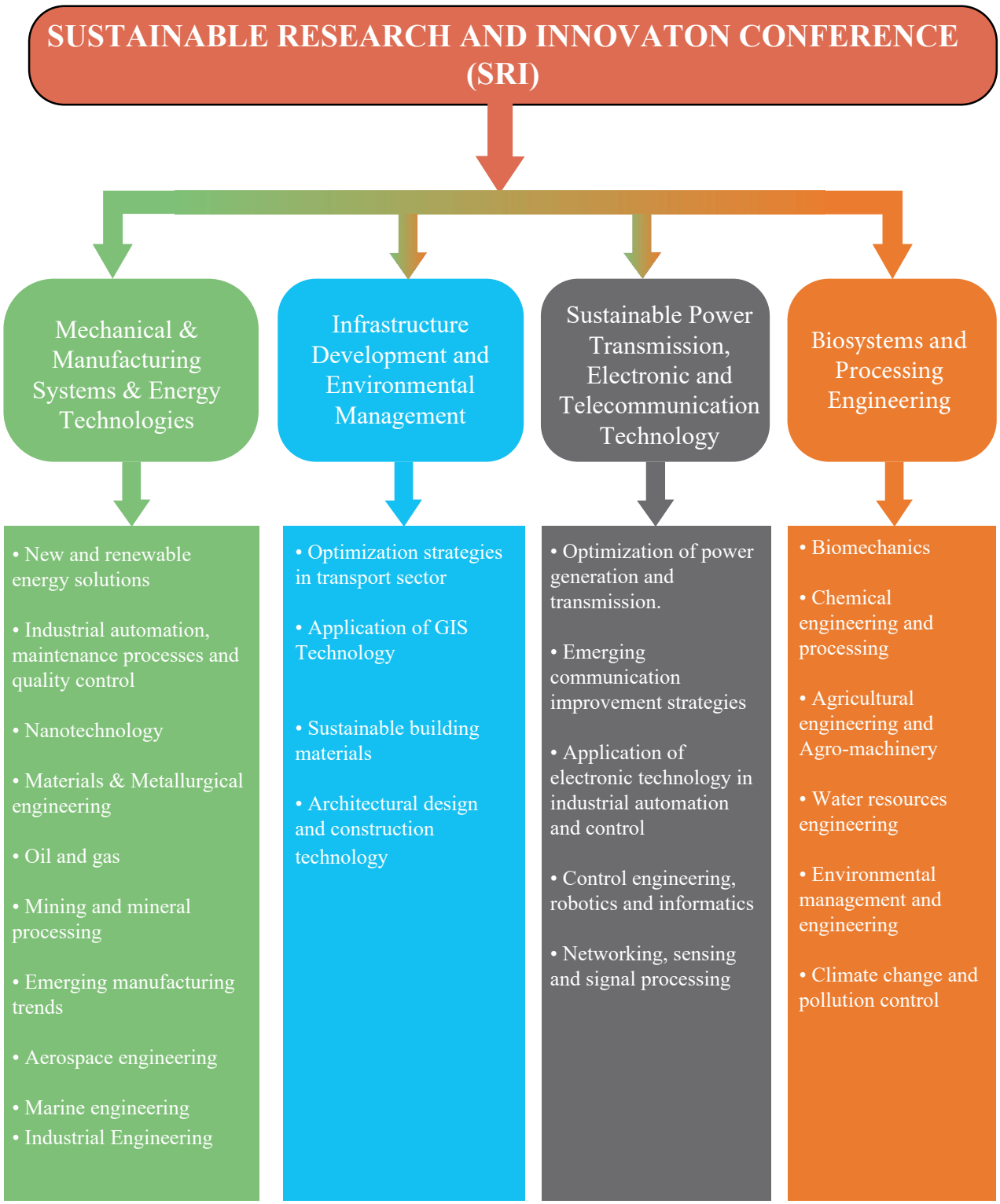
Research, in the 21st century, is dependent on the multi-disciplinary approach for development of new knowledge and products. Industrialized countries spend a significant percentage of their GDP on funding research in universities and other dedicated research institutions and centres. The challenge for developing countries, however, is that of building up capacity in order to ensure that industries remain competitive in the global markets, in addition to offering innovative solutions to the increasing socio-economic problems and developing efficient methods of using the depleting natural resources. The multi-disciplinary approach is a possible method of tackling these challenges. The conference offers a wide selection of topics to give the participants an opportunity to share experiences and link research to the process of industrialization.

The theme for the 2018 Annual Sustainable Research and Innovation Conference is "Emerging Technological Innovations for Sustainable Industrial Development". The conference aims to create a forum for scholars, industry and other stakeholder to interact and exchange ideas on new and existing knowledge in the field of engineering and related subjects. The conference is also a forum for enabling participants to articulate challenges facing the society, and offer viable solutions. In addition, scholars are also able to create networks for multi-disciplinary and industrial based research that will solve national and international challenges. The wide selection of topics listed herein give the participants an opportunity to share experiences and articulate how their activities impact on the process of industrialization in Africa and other developing countries.

The Sustainable Research and Innovation (SRI) Conferences is the successor to the Departmental Annual Seminars which were previously organized by the Department of Mechanical Engineering at JKUAT. The first seminar was held in 1995, under the theme: "The Role of Mechanical Engineering in Changing Industry". Thereafter, seven seminars were held annually until 2001. After a few years' break, the annual seminars resumed in 2006 with the 8th series titled "Sustainable Research and Innovation". In 2010, these Departmental Seminars were scaled to international conferences to run annually organized by the School of Mechanical, Manufacturing and Materials Engineering up to the year 2014 after which the role of organnizing this conference was taken up by the college of Engineering. The SRI conferences have offered a platform for researchers and innovators in industry and academia to disseminate their research findings over the years.

This year's, conference has drawn participation from the academic field and industries in Kenya as well as various countries including Japan, Zimbabwe, South Africa, Botswana Ghana, Rwanda and Germany. The conference has been supported by Jomo Kenyatta University of Agriculture & Technology (JKUAT), Japan International Cooperation Agency (JICA) through Africa-ai-Japan Project and Kenya's National Research Fund (NRF).

The papers presented during the 2018 Sustainable Research and Innovation Conference were based on the following topics



ACKNOWLEDGEMENTS

The SRI conference seeks support from well-wishers in order to ensure that all the activities in preparation of and during the conference run smoothly. The donors have no influence over the selection or content of the presentations.

In 2018, the conference has received financial support commitment from Jomo Kenyatta University of Agriculture and Technology – JKUAT, Japan International Cooperation Agency – JICA through the Africa-*ai*-JAPAN project and National Research Fund – NRF.

Platinum

Platinum

Silver

THE ADVISORY BOARD

Prof. Eng. B.W. Ikua
Jomo Kenyatta University of Agriculture and Technology, Kenya

Prof. G. Nyakoe
Jomo Kenyatta University of Agriculture and Technology, Kenya

Prof. S.M. Maranga
Jomo Kenyatta University of Agriculture and Technology, Kenya

Prof. Eng. J.M. Kihiu
Jomo Kenyatta University of Agriculture and Technology, Kenya

Prof. Eng. S.P. Ng'ang'a
Jomo Kenyatta University of Agriculture and Technology, Kenya

Dr. K. K. Kaberere
Jomo Kenyatta University of Agriculture and Technology Kenya

Prof. B. M. Mati
Jomo Kenyatta University of Agriculture and Technology Kenya

Prof. P. N. Kioni
Jomo Kenyatta University of Agriculture and Technology, Kenya

Dr. Eng. K. Kibicho
Interior and Coordination of National Government, Kenya

Prof. F. Gatheri
Technical University of Kenya

Prof. G. O. Rading
University of Nairobi, Kenya

Prof. E. Sciubba
University of Roma, Italy

Prof. E. Tomita
Okayama University, Japan

Prof. H. Tanaka
Tottori University, Japan

Prof. L. Zuyan
Harbin Institute of Technology, P.R. China

Prof. K. Badyda
Warsaw University of Technology, Poland

Prof. J. Milewski
Warsaw University of Technology, Poland

Prof. A. Beyene
San Diego State University, USA

Prof. Eng. Bernard Ikua
Principal, College of Engineering and Technology

Dr. Eng. Hiram Ndiritu
Dean School of Mechanical, Mechatronics and Materials Engineering,

Dr. Eng. Gareth Kituu
Dean, School of Biosystems and Environmental Engineering

Prof. Abiero Gariy
Dean, School of Civil, Environmental and Geomatic Engineering

Prof. John Nderu
Dean, School of Electrical, Electronics and Information Engineering

Dr. Onesmus Muvengei
Chairman, Mechanical Engineering Department

Dr.-Ing Jackson G. Njiri
Chairman, Mechatronic Engineering Department

Dr. Julius Weru
Chairman, Telecommunication and Information Engineering Department

Prof. Stanley Kamau
Chairman, Electrical and Electronic Engineering Department

Prof. Bernard K. Rop
Chairman, Mining, Materials and Petroleum Engineering Department

Dr. Eng. Jeremiah Kiptala
Chairman, Civil Engineering Department

Dr. Felix Mutua
Chairman, GEGIS Department

Dr. Stephen Ondimu
Chairman, Agricultural and Bio-Systems Engineering

Dr. Eng. James Messo
Chairman, Soil Water and Environmental Engineering Department

Dr.-Ing Jackson G. Njiri
Dept. of Mechatronics Engineering
Chair, Secretariat

Mr. Peter Mwangi
Dept. of Mechanical Engineering
Treasurer

Mr. Shem Otoi
Dept. Marine Engineering & Maritime Operations
Secretary

Mr. Gilbert Bett
Dept. Mining, Materials, and Petroleum Engineering
Communications Liaison

Ms. Laura Simiyu
Dept. Mining, Materials & Petroleum Engineering
Member

Ms. Mercy Ndinda
Dept. of Electrical and Electronic Engineering
Member

Ms. Joan Maina
Dept. of Agricultural and Bio-Systems Engineering.
Member

Mr. Linus Aloo
Dept. of Electrical and Electronic Engineering
Member

Mr. Halake Guyo
Dept. of Soil, Water and Environmental Engineering
Member

Ms. Dorcas Chege
Dept. of Mechanical Engineering
Member

Ms. Irene. Fedha
Dept. Mining, Materials & Petroleum Engineering
Member

Mr. Omae Oteri
Dept. of Telecommunication and Information Engineering (TIE)
Member

TABLE OF CONTENTS

Utilization of FACTS devices in power systems: A review

Irene N. Muisyo and Keren K. Kaberere

Abstract— Electricity demand has been increasing over the last three decades, while the expansion of generation and transmission networks has been limited due to environmental and economic constraints of building new generating plants and transmission lines. Consequently, transmission networks are at times driven close to their thermal and stability limits. Over the years, attention has been shifting towards better utilization of existing power system infrastructure, which can be achieved through employment of power electronic technologies such as flexible AC transmission system (FACTS) devices. FACTS devices have been used to solve various power system problems such as voltage regulation, power flow control, transfer capability enhancement, damping inter area modes and enhancing power system stability. This paper presents a literature survey of FACTS devices and their applications in power systems.

Keywords— FACTS, SVC, TCSC, STATCOM, SSSC, UPFC.

I. Introduction

THE power system is the most complex system built by man. It consists of numerous components such as generators, transmission lines, distribution lines, transformers and a variety of loads. The efficiency of a power system is determined by the ability of transmission lines to optimally transfer electrical power from generating stations to the loads. Power flow control in power systems is increasingly becoming a major concern for system operators as a result of power system restructuring and constrained transmission expansion [1]. Conventional power flow control techniques such as generation rescheduling, use of phase shifting transformers, reactive power support and load shedding are frequently applied in power systems.

In deregulated markets, generation rescheduling and load shedding have to be limited since there are existing power transaction contracts between generating companies and their customers. Reactive power support and the use of FACTS devices to enhance power system performance has gained a lot of interest in recent years [2] [3]. This paper presents an overview of FACTS devices, their applications in power systems and their associated costs and benefits. The paper is organized as follows: in Section 2, an introduction to FACTS devices is given whereas the different types of FACTS devices is presented in Section 3. FACTS device applications are outlined in Section 4 while their benefits and costs are given in Section 5. The emerging research areas related to FACTS device utilization is given in Section 6. Conclusions are drawn in Section 7.

II. Flexible AC Transmission Systems

The concept of FACTS was first introduced by Hingorani and Gyugyi in 1988. FACTS refer to the application of power semi-conductor devices to control electrical variables such as voltage, impedance, phase angle, current, active and reactive power [1]. The active power transferred in a lossless line from bus *i* to bus *j* is dependent on line reactance X_{ij}, bus voltage magnitudes V_i and V_j and the phase angle between sending and receiving end buses δ, as approximated in (1).

$$P_{ij} \approx \frac{V_i V_j}{X_{ij}} \sin \delta \qquad (1)$$

FACTS devices act by modifying the line reactance, injecting or absorbing reactive power (thus influencing bus voltage magnitude) or modifying the phase angle between sending and receiving end buses. FACTS devices are designed using high speed power electronic controllers. This overcomes limitations of mechanically controlled AC power transmission system components which include slow response and wear and tear [3].

Increased utilization of FACTS devices has been made possible due to the ongoing developments in the field of power electronics. In practical systems, some lines become congested if unplanned power exchanges occur. To enhance the security of the power system, appropriate types of FACTS devices should be chosen and suitably installed to redistribute power flow and regulate bus voltages [2] [4].

The optimal location of FACTS devices is a highly constrained and complex optimization problem. Generally, the approaches towards optimal location of FACTS devices can be classified into three categories: classical optimization approaches, sensitivity based approaches and heuristic optimization techniques [5]. FACTS devices can be connected to a transmission line at any appropriate location in series, shunt or a combination of series and shunt. Shunt FACTS controllers include the static VAr compensator (SVC) and static synchronous compensator (STATCOM), series FACTS controllers include the thyristor controlled series compensator (TCSC) and static synchronous series compensator (SSSC) whereas the combined shunt and series FACTS controllers include unified power flow compensator (UPFC) and interline power flow compensator (IPFC) [6] [7].

I. N. Muisyo, Department of Electrical and Electronic Engineering, JKUAT (Mobile Phone: +254721332377; e-mail: muisyoirene@jkuat.ac.ke).

K. K. Kaberere, Department of Electrical and Electronic Engineering, JKUAT (e-mail: kkanuthu@eng. jkuat.ac.ke).

From literature survey, various researchers have investigated utilization of FACTS devices for different objectives such as maximizing power system security [2] [8], enhancing power system transient stability [9] [10], enhancing voltage stability [11], minimizing active power losses [12], minimizing cost of generation [13] among others. Different metaheuristic optimization techniques such as cuckoo search algorithm (CSA) [2], particle swarm optimization (PSO) [14], non-dominated sorting genetic algorithm (NDSGA) [15], improved particle swarm optimization algorithm using eagle strategy (ESPSO) [12], evolutionary algorithm [11], genetic algorithm [13] have been investigated. Classical optimization techniques such as nonlinear programming methods [3] and line flow indices [8] have also been explored. Multi-objective optimization problems for FACTS device location have emerged in the last decade, as outlined in [14] [15]. Emerging areas of FACTS device utilization in power systems have also been explored such as the integration of renewable energy sources (RES) and smart grids [16], and parameter tuning of power system stabilizers (PSS) [17] [18]. We now describe various FACTS devices and their utilization in power systems.

III. CLASSIFICATION OF FACTS DEVICES

There are three generations of power electronic based FACTS devices:

i. First generation FACTS devices which employ thyristor switched capacitor and reactor banks to control on and off periods of the devices.
ii. Second generation FACTS devices which use self-commutated DC-AC converters to generate capacitive and inductive reactive power for transmission line compensation.
iii. Last generation FACTS devices also known as D-FACTS which are smaller and less expensive than first and second generation FACTS devices. They are used in distribution systems [16].

The basic principles of operation of FACTS devices will be discussed briefly.

A. First generation FACTS devices

The first generation FACTS devices are: static VAr compensator (SVC), thyristor controlled series capacitor (TCSC), and thyristor controlled phase shifter (TCPS).

1) Static VAr compensator (SVC)

The SVC is the oldest FACTS device whose primary purpose is to improve bus voltages by means of reactive power compensation. The world's first SVC for utility application was installed in 1974 in Nebraska by General Electric (GE). More than 800 SVCs have been installed worldwide, both for utility and industrial applications. ABB remains the pioneer in development of SVC and has supplied 55% of the total market of which 13% were being installed in Asian countries [16]. As a consequence of deregulation in the UK in 1990, voltage control became difficult. To accommodate the risk associated with changing power system conditions, the UK installed relocatable SVCs in the National Grid Company (NGC). By 2004, 12 RSVCs of 60 MVAr each were operational [9]. The basic circuit of SVC is shown in Fig. 1.

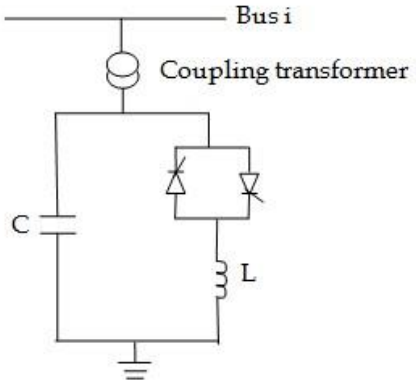


Fig. 1 Basic circuit for SVC

As shown in Fig. 1, the SVC is composed of a fixed capacitor (C) and a thyristor controlled reactor (L). The SVC is usually connected at the midpoint or at the end of a transmission line through a coupling transformer. The equivalent susceptance of the SVC is controlled by adjusting the firing angle of the thyristors. With proper coordination of the capacitor switching and reactor control, the VAr output can be varied continuously between the capacitive and inductive ratings of the equipment. SVCs are used for steady state voltage control, damping oscillations, improving transient stability and reducing system losses [17].

2) Thyristor controlled series compensator (TCSC)

The TCSC is a first generation FACTS device which uses thyristors to manage a capacitor bank connected in series with a transmission line. The world's first three phase TCSC was developed by ABB and installed at Kayenta substation, Arizona in 1992, that raised the capacity of a transmission line by almost 30%. By 2004, 7 TCSC's had been installed around the world such as in Sweden, China and India [16]. The TCSC consists of a capacitor bank C, and a thyristor controlled reactor (L) as shown in Fig. 2.

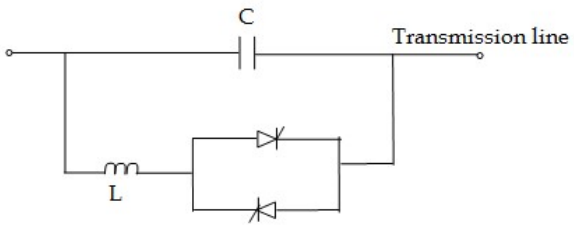


Fig. 2 Basic structure of TCSC

The TCSC is connected in series with the AC transmission line whose capacity of power transmission is to be improved. By adjusting the firing angle of the thyristors, the capacitive reactance is smoothly controlled over a wide range, thus the line

reactance is modified. The TCSC is increasingly finding use in long transmission lines for functions such as increasing line loadability, mitigating sub-synchronous resonance, damping power oscillations and enhancing transient stability [13] [19].

B. Second generation FACTS devices

The second generation FACTS devices are: static synchronous compensator (STATCOM), static synchronous series compensator (SSSC), unified power flow controller (UPFC) and interline power flow controller (IPFC).

1) Static synchronous compensator (STATCOM)

The STATCOM is a gate turn off (GTO) based voltage source converter (VSC) capable of generating or absorbing real and reactive power at its output terminals. The world's first commercial STATCOM (80 MVA, 154 kV) was installed in Inuyama substation in Japan in 1991 by Mitsubishi Electric Power Products Inc. A STATCOM with a capacity of 225 MVAr was also established in East Claydon 400kV substation, UK in 2001, whose purpose was to provide dynamic reactive compensation. Another STATCOM of 100 MVAr was constructed in SDG & E Talega substation, USA in 2003, with the aim of providing dynamic VAr control during peak load conditions. In 2013, a STATCOM with capacity of 150MVAr at 275kV was installed in Turkey too. There are more than 20 STATCOMs operating successfully around the world [16] [18]. The basic structure of a STATCOM is shown in Fig. 3.

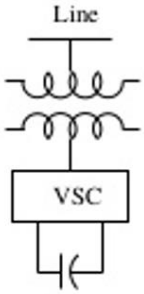


Fig 3 Basic structure of STATCOM

As shown in Fig. 3, the STATCOM is a shunt connected static VAr compensator which converts DC input voltage into AC output voltage in order to compensate the active and reactive power needed by the system. The STATCOM is able to control its output voltage independently of the AC system voltage, hence can provide better reactive power compensation at low grid voltages [9] [18].

2) Static synchronous series compensator (SSSC)

The SSSC is made up of a voltage source converter (VSC) serially connected to a transmission line through a transformer. The SSSC is not yet in commercial operation as an independent controller. Its basic structure is shown in Fig. 4.

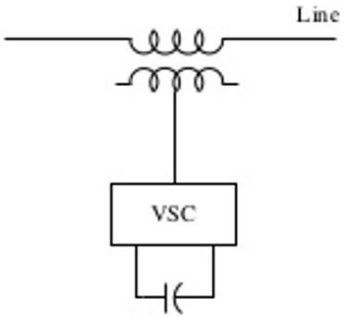


Fig. 4 Basic structure of SSSC

From Fig. 4, the SSSC is based on a DC capacitor fed VSC which generates a three phase voltage at fundamental frequency, which is then injected in a transmission line through the series transformer. The SSSC injects a voltage with controllable magnitude and phase angle at the line frequency, hence it controls the active and reactive power flow in the network. The main advantage of SSSC over TCSC is that it does not remarkably interfere with the transmission line impedance hence there is no danger of having resonance problems [10].

3) Unified power flow controller (UPFC)

The SSSC and a STATCOM can be combined to produce a unified power flow controller (UPFC). The first utility UPFC was installed at the Inez substation of American Electric Power in 1998. In 2004, an 80 MVA UPFC was also constructed at Gangjin substation in South Korea. The basic structure of UPFC is shown in Fig. 5.

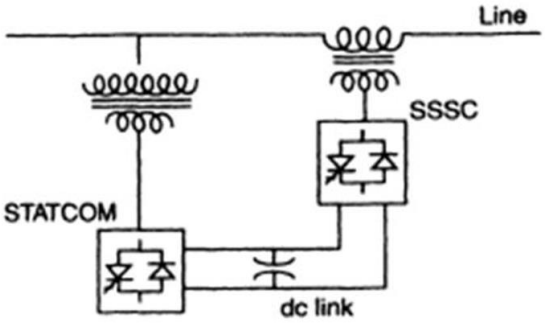


Fig. 5 Basic structure of UPFC

From Fig. 5, the UPFC consists of two AC/DC converters, one connected in series with the transmission line and the other in parallel with the transmission line. The DC side of the two converters is connected through a common capacitor which provides DC voltage for the converter operation. The series converter injects an AC voltage with controllable magnitude and phase angle in series with the transmission line via a series transformer. The shunt converter injects or absorbs the real power demand of the series converter at the common DC link. It also generates or absorbs reactive power to provide shunt compensation for the line. The UPFC thus allows exchange of real and reactive power with the transmission line, hence improving steady state and transient stability of the system [20] [21].

C. D-FACTS devices

A new concept of smaller and less expensive FACTS devices is being explored, known as distributed FACTS devices. The adoption of D-FACTS will enhance performance of power systems at a lower installation cost. D-FACTS will also be used to solve some challenges arising in smart grids and micro grids such as voltage fluctuations and interruptions. They will enhance controllability, reliability and improve end user power quality with little environmental impact. D-FACTS devices include D-STATCOM, D-SSSC and D-UPFC [16] [22].

Next we outline some FACTS device applications.

IV. FACTS DEVICE APPLICATIONS

FACTS devices find applications in all three states of the power system namely: steady state, transient and post transient steady state.

A. Steady state applications of FACTS devices

The steady state applications of FACTS devices include:

i. Congestion management usually carried out by the system operator. FACTS devices such as TCSC, TCPAR and UPFC are employed to alleviate congestion.

ii. Available transfer capacity (ATC) improvement. Low ATC implies that the network cannot accommodate new transactions. FACTS controllers such as TCSC, TCPAR and UPFC can be employed to increase the ATC.

iii. Reactive power and voltage controllers like SVC and STATCOM are used for voltage control.

iv. Loading margin improvement. Both series and shunt compensators are used to increase the transfer capabilities of power systems [13] [23].

B. Dynamic applications of FACTS devices

Dynamic applications of FACTS devices include:

i. Transient stability enhancement. FACTS devices can provide rapid response during the first swing to control the voltage and power flow in the system.

ii. Oscillation damping. Electromechanical oscillations can be better damped if power system stabilizers (PSS) are coordinated with FACTS devices.

iii. SSR mitigation. Sub-synchronous resonance (SSR) associated with conventional series capacitors can be mitigated by controlling the operating modes of series FACTS devices such as used in Stode, Sweden power system.

iv. Power system interconnection. With series compensation, bulk AC power transmission over distances of more than 1,000km are a reality today as has been used in Brazil North – South interconnection [16] [17].

C. Application in deregulated markets

FACTS controllers are finding new applications in deregulated markets. Some of these applications include:

i. Controlling loop flows which can reduce the transmission capacity, and hence limit the possible transactions through a given line.

ii. Making use of FACTS controllers in tie lines to enable participation in transactions or to shield them from the neighboring wheeling transactions.

iii. Optimally locating FACTS devices to ensure economic dispatch and reduction in power system losses [23].

The benefits and costs of utilizing FACTS devices in power systems are presented below.

V. BENEFITS AND COSTS OF FACTS DEVICES

The investment costs of FACTS devices are huge, as seen from the world's second UPFC, which installed in the year 2004 in Keepco power system, Korea. This was the largest single procurement order ever placed by Keepco power system [16]. The investment costs of FACTS devices therefore have to be computed against anticipated benefits.

A. Benefits of FACTS device utilization

The benefits of utilizing FACTS devices in power systems are:

i. Environmental benefits

FACTS devices are installed on existing infrastructure, hence eliminate the issue of encroaching on new public and private land. For example in Sweden, eight 400kV systems run parallel to transmit electricity from the north to the south. Each of these transmission lines are equipped with FACTS devices. Studies have shown that four additional 400kV transmission lines would have been necessary, if FACTS devices had not been installed on the power system [16].

ii. Enhanced transmission system reliability

FACTS devices employ power electronic devices to regulate power flow and transmission network voltages through fast control action. This can mitigate dynamic disturbances thus increasing system reliability.

iii. Rapidly implemented installations

Constructing a new transmission line may take several years, whereas FACTS device installation takes between 12 to 18 months. FACTS device installation also allows the flexibility for future upgrades.

iv.Financial benefits

Financial benefits from FACTS device utilization comes from additional sales or wheeling charges due to increased transmission capability, delay in investing in high voltage transmission lines and engaging cheaper generation facilities. The enhanced power system stability also reduces the risk of forced outages, thus reducing lost revenue and penalties from power contracts.

v. Reduced maintenance cost

The overhead transmission lines need to be cleared from the surrounding environment from time to time. In comparison to this, the FACTS maintenance cost is very minimum [16] [23].

B. Drawbacks of FACTS technology

The major limitation of FACTS device utilization is their high cost of investment. The investment costs mostly depend on the

device rating and modifications to be done on existing infrastructure. The higher the device rating, the more expensive the device is. Due to the modifications to be done on existing substations, costs such as civil works, installation, commissioning, insurance and project management will arise. The resulting voltage and current waveforms from FACTS devices are usually distorted due to the switching nature of power electronic converters. Additional interface filters have to be incorporated [5] [6].

VI. RESEARCH GAPS

There is a lot of ongoing research in terms of FACTS device utilization ranging from their technologies, optimal location, coordination of FACTS devices with power system components and the economic benefits of FACTS device utilization.

A. FACTS device technologies

With the history of more than three decades and widespread research and development, FACTS controllers are now considered a proven and mature technology. From various studies, the performance of second generation FACTS devices is seen to be better than for first generation devices, with the UPFC termed as the most versatile FACTS device [4] [8] [9] [10]. From literature survey, very little has been done on the actual AC/DC models of VSC FACTS devices. An accurate representation of VSC FACTS devices in transient stability studies would be beneficial to bring out their behavior in stressed system conditions, or when the devices operate at their limits. The last generation of VSC based D-FACTS devices should be thoroughly investigated to depict how their topologies can be used to enhance performance of micro-grids, standalone DG schemes, RES integration and smart grids [2] [5] [16].

B. Optimal location of FACTS devices

Determining the best location and size of FACTS devices in a highly interconnected network is also a complex task. Evolutionary computation and random search algorithms such as GA, PSO, SA, CSO techniques have all been proposed to find the optimal location of FACTS devices. These techniques have some advantages especially when dealing with non-differential and non-convex problems, but have poor scalability and repeatability. Sensitivity based approaches such as weighted indices are frequently used but cannot ensure optimality. To complement the shortcomings of different optimization techniques, hybridization, especially of metaheuristic algorithms, should be adopted [2] [3]. To simultaneously solve power system problems such as reactive power dispatch, cost minimization, loss minimization, stability enhancement using FACTS devices, multi-objective problem formulation should be adopted [11]- [15].

C. Coordination of FACTS devices with power system components

As the number of FACTS devices in a power system increases, interactions among themselves will be a serious concern. Interactions can also take place between the FACTS devices and PSS, HVDC systems or any other system controllers. Therefore, coordination among various FACTS devices and power system components should be investigated in detail, both in steady state and in the event of a disturbance [17] [18].

D. Economic benefits of FACTS device utilization

Lastly, very few researchers have addressed the economic benefits of utilizing FACTS devices. Studies should be carried out on existing FACTS projects in terms of installation, maintenance and other associated costs of FACTS device utilization. Costs should be weighed against the actual benefits in terms of wheeling charges, savings on cost of generation, savings due to improved system reliability among others [2] [3].

VII. CONCLUSION

This paper gives a review of various FACTS devices and their application in power systems. The benefits and costs associated with FACTS devices, and research gaps in the field of FACTS device utilization are outlined. We observe that accurate models of VSC FACTS devices should be studied in depth, with focus on the last generation FACTS device technologies. In simultaneously solving power system problems, multi-objective optimization problems should be solved using hybridized metaheuristic techniques. Another research gap lies in how FACTS devices interact with other power system components and how their controllers can be simultaneously coordinated for dynamic stability enhancement. Finally we conclude that the actual costs and benefits of FACTS device utilization should be thoroughly investigated, especially on existing projects.

References

[1] A. Peter, A. Anthony, A. Claudis and A. Abel, "A review of the applications of FACTS devices on Nigerian 330kV transmission system," *Journal of Engineering and applied sciences,* vol. XII, no. 20, pp. 5182-5185, 2017.

[2] T. Kang, J. Yao, T. Duong, S. Yang and X. Zhu, "A hybrid approach for power system security enhancement via optimal installation of flexible AC transmission system (FACTS) devices," *Energies,* vol. I, no. 10, pp. 1-32, 2017.

[3] X. Zhang, K. Tomsovic and A. Dimitrovski, "Optimal investment on series FACTS device considering contingencies," National Science Foundation, 2017.

[4] R. Sikka, R. Bahl and V. Verma, "Review on comparison of FACTS controllers for power system stability enhancement," *International Interdisciplinary Conference on Science Technology Engineering Management Pharmacy and Humanities,* vol. I, no. 1, pp. 316-320, 2017.

[5] N. Khatoon and S. Shaik, "A survey on different types of flexible AC transmission systems (FACTS)

controllers," *International Journal of Engineering Development and Research*, vol. IV, no. 5, pp. 796-814, 2017.

[6] G. Barve, "Application study of FACTS devices in Indian power system," *International Journal of Computing and Technology*, vol. I, no. 1, pp. 57-60, 2014 .

[7] M. Singh, "Power flow with flexible alternating current transmission systems (FACTS) controller," *IJARIIT*, vol. 1, no. 1, pp. 180-184, 2018 .

[8] I. Khan, M. A. Mallick, M. Rafi and M. S. Mizra, "Optimal placement of FACTS controller scheme for enhancement of power system security in Indian scenario," *Journal of Electrical Systems and Information Technology*, vol. 1, no. 2, pp. 161-171, 2015.

[9] M. Verma and S. Jain, "Improving power system transient stability by using FACTS devices," *International Journal for Technological research in Engineering*, vol. IV, no. 9, pp. 1579-1582, 2017 .

[10] "Improvement of Power System Transient Stability using TCSC, SSSC and UPFC Internatinal Journal of Scientific & Engineering Research Lokesh Garg; S. K. Agarwal; Viviek Kumar," vol. 8, no. 4, pp. 57-61, 2017.

[11] S. d. Nascimento and M. M. G. Jr, "Voltage stability enhancement in power systems with automatic FACTS device allocation," *International Conference on Energy and Environment Research*, vol. 1, no. 107, pp. 60-67 , 2017.

[12] H. Yapici and N. Cetikaya, "An Improved Particle Swarm Optimization Algorithm using Eagle Strategy for Power Loss Minimization," *Mathematical Problems in Engineering*, vol. 1, no. 1, pp. 1-12, 2017.

[13] C. H. B. Apribowo, M. H. Ibrahim and F. X. R. Wicaksono, "Optimal power flow with optimal placement TCSC device on 599kV Java-Bali electrical power system using Genetic Algorithm-Taguchi method," *AIP conference proceedings*, vol. 1, no. 1, pp. 1-10, 2018.

[14] P. Ramesh, A. Chiranjeevi and K. Padma, "Application of SVC for Multi-objective Optimization Powerflow Problem using PSO," *International Journal of Pure and Applied Mathematics*, vol. 114, no. 8, pp. 23-33, 2017.

[15] M. Nafar and A. Ramezanpour, "Optimal Allocation of TCSC using Heuristic Optimization Technique," *International Journal of Scientific Study*, vol. V, no. 5, pp. 250-254, 2017.

[16] F. H. Gandoman, A. Ahmad, A. M. Sharaf, P. Siano, J. Pou, B. Hredzak and V. Agelidis, "Review of FACTS technologies and applications for power quality in smart grids with renewable energy systems," *Renewable and sustainable energy reviews*, vol. 1, no. 82, pp. 502-514, 2018.

[17] M. O. Benaissa, S. Hadjeri and S. A. Zidi, "Impact of PSS and SVC on the power system transient stability," *Advances in Science, Technology and Engineering Systems Journal*, vol. II, no. 3, pp. 562-568 , 2017.

[18] K. Karthikeyan and P. Dhal, "Optimal location of STATCOM based dynamic stability analysis tuning of PSS using partice swarm optimization," *International conference on processing of Materials, Minerals and Energy*, vol. 1, no. 5, pp. 588-595, 2018.

[19] A. R. Kumbhaj, "A review on power stability improvement using FACTS controllers in power systems," *International Journal of Industrial Electronics and Electrical Engineering* , vol. V, no. 11, pp. 10-14, 2017.

[20] G. B. Jadahav, C. B. Bangal and S. Kanungo, "Transient stability analysis with SVC and STATCOM in multi-machine power systems with and without PSS using Matlab/Simulink," *International Journal of Engineering Development and Research*, vol. V, no. 4, pp. 1196-1205 , 2017.

[21] S. V. Patil and K. Mahajan, "A review on implementation of UPFC for improvement of active power flow capability in power system using IEEE 14 bus system," *International Research Journal of Engineering and Technology (IRJET)*, vol. IV, no. 4, pp. 542-547, 2017.

[22] P. Tipathi and G. P. Pandya, "A survey on the impact of FACTS controllers on power system performance," *International Journal of Engineering Trends and Technology (IJETT)*, vol. 64, no. 1, pp. 24-29, 2017.

[23] D. Kothari and I. J. Nagrath, Modern power system analysis 3rd edition, New Delhi: Tata McGraw Hill Education Private Limited , 2003.

[24] I. Khan, M. A. Mallick, M. Rafi and M. S. Mizra, "Optimal placement of FACTS controller scheme for enhancement of power system security in Indian scenario," *Journal of Electrical Systems and Information Technology*, no. 2 , pp. 161-171, 2015.

[25] A. Khasdeo, "Transient stability improvement in transmission system using SVC with fuzzy logic control," *International Research Journal of Engineering and Technology*, vol. IV, no. 4, pp. 173-177, 2017.

[26] N. Khan, H. Ahmad and A. Khattak, "Transient stability enhancement of power system using UPFC (unified power flow controller)," *International Journal of Engineering Works*, vol. IV, no. 2, pp. 33-40, 2017.

A Review of Approaches to the Quantification of Operating Reserve Requirements in Systems with Large Scale Wind Power Integration

Brian K. Wamukoya, Bonface O. Ngoko, and Christopher M. Muriithi

Abstract—The need to diversify power generation sources, increase security of supply, incorporate sustainability in energy sources and reducing fuel usage and emissions has significantly led to the integration of variable renewable energy sources such as wind to power grids globally. However, increased integration of wind power into a grid necessitates the need to update unit commitment and operating reserve algorithms since wind power is highly variable, intermittent and non-dispatchable. In addition, operational decisions in these grids are made on the basis of wind and load forecasts which are not perfect. Therefore, variability and uncertainty due to wind power introduced in the system requires an increase in operating reserve allocation to maintain reliability levels after wind power integration. This paper reviews recent work on determination of additional operating reserve requirement for power systems with large scale wind power integration. It also highlights different methods incorporated in solving the optimization problem in balancing between economics and reliability of power systems in sizing of additional reserve. In conclusion, gaps in literature that require further research are identified.

Keywords—Large scale wind integration, Operating reserve, Reserve quantification, Wind variability and uncertainty.

I. INTRODUCTION

IN the last couple of decades, the need to diversify power generation sources, increase security of supply, incorporating sustainability in energy sources and reducing fuel usage and emissions has significantly led to the integration of variable renewable energy sources (VREs) such as Wind power to power grids globally. Integrating VREs is far much less complicated if they are incorporated to large power systems which can take advantage of natural diversity of variable sources [1]. However as wind capacity increases, considerable measures have to be taken to ensure that the wind power variation does not reduce the reliability of the power system.

Wind power possess great challenges to power system operators in grid management and generation scheduling as a sudden and severe loss of generation might occur at a high penetration level of wind power. The inherent variability, intermittency and uncertainty that characterize wind power require that current industry practices be altered which necessitates a need to update unit commitment and operating reserve procurement algorithms to accommodate smooth integration [1], [2]. The additional variability and uncertainty introduced in the system will result in an increase in operating reserve requirement in the system to maintain reliability levels after wind integration [3]–[5]. The objective is to meet high load demand and withstand the impact of wind power fluctuations.

The requirement for additional reserve capacity is determined by critically monitoring the variation of wind power production, hourly and intra-hourly, together with load variations and prediction errors [1]. Therefore accurate forecasting tools such as numerical weather prediction are critical to wind integration as it contributes to risk reduction. Accuracy of wind power production forecasts depends on several factors such as the forecast horizon, the size of wind power plants and their geographical dispersion, experience with wind power generation and the accuracy of forecasts for individual wind power plants [6]. Wind power forecast errors (WFEs) increase as the forecasting horizon gets longer [6]–[8]. Large geographical spreading of wind power will also increase predictability [9], [10].

Since the cost associated with provision of additional reserve capacity as a result of wind integration is far from being negligible, quantification of operating reserve has become a subject of intense interest for researchers and power engineers in the recent past. In this regard, system models and analytical methods using time series of wind power production together with a host of other power system variables and constraints have previously been presented to estimate reserve requirement and the cost associated with it [1].

Statistical methods that rely on the use of standard deviation (σ) like the n-sigma criterion are one of the most widely used to quantify the effects of increased wind power on operating reserves due to its simplicity. However, this method is most applicable in power systems where the wind power capacity is decentralized over a large geographical area with different wind flow regimes. Consequently, the variability of total wind power injections is attenuated, even more with increasing number of wind turbines (WTs) installed at different locations [11], [12]. Under these circumstances, both wind power variability and wind forecast errors (WFE) are assumed to follow the Gaussian distribution [13], [14]. Suffice to mention, wind variability and forecast errors do not necessarily follow a Gaussian distribution [15], [16].

Alternatively, sophisticated probabilistic methods can be deployed to quantify operating reserve by integrating different sources of uncertainties using reliability theory of power systems [17]. They are based on generation margin which is defined as the difference between total available generation

B. K. Wamukoya, PAUWES, Pan African University (corresponding author phone: +254 715880215; e-mail: brianwamukoya@gmail.com).

B. O. Ngoko, Division of Electrical, Electronic, and Information Engineering, Osaka University (e-mail: ngokobonface@gmail.com).

C. M. Muriithi, Department of Electrical and Electronic Engineering, Murang'a University of Technology (e-mail: cmmuriithi@gmail.com).

and load. Classical reliability indices such as loss of load expectation (LOLE), loss of load probability (LOLP) and expected energy not supplied (EENS) are calculated based on the distribution of generation margin and used in the sizing of appropriate operating reserve levels.

This paper summarizes recent work on the determination of operating reserve requirement for power systems with large scale wind integration with the main focus on solution approaches on the sizing and allocation of different categories of operating reserves. It also highlights the different methods incorporated in solving the optimization problem in the balancing between economics and reliability of the power system in the sizing and allocation of operating reserve requirement. Eventually, research gaps are identified after a thorough evaluation of existing methodologies in this field of research.

II. PROBLEM DESCRIPTION

Problem formulation in quantifying additional reserve requirement is mostly anchored on balancing minimization of operating cost with guaranteeing required reliability levels while considering other system constraints.

A. Objective Function

The objective is not to be overly conservative in scheduling additional operating reserve racking up costs unnecessarily while also not excessively minimize cost to levels compromising system security. The objective function of the reserve allocation problem in [18] is given as:

$$\min \sum_{i=1}^{N_R} \rho_i R_i \tag{1}$$

where N_R is the number of generators to provide operation reserve, R_i- is the selected reserve of the generator, ρ_i- is the reserve cost of generator i. An illustration of the cost (reserve) / reliability curve (EENS) in [19] is shown in Fig. 1.

Reference [18] focuses on the day-ahead allocation of operation reserve considering wind power forecast error and network transmission constraints in a composite power system.

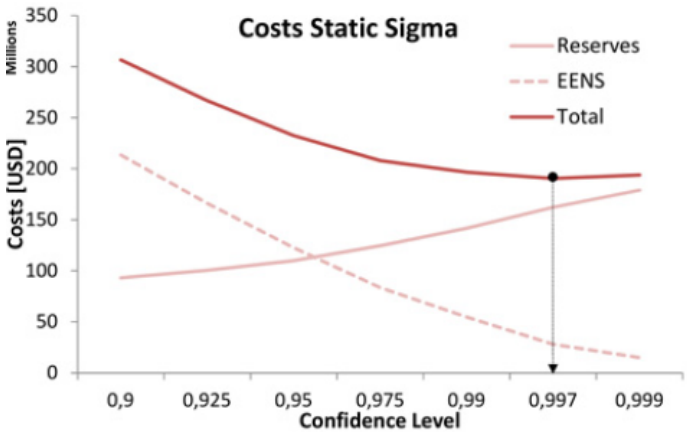


Fig. 1. Cost vs. Reliability curves [19]

The authors in their methodology were determined to minimize operating cost associated with allocation of operating reserve while at the same time meeting required system reliability levels.

Reference [19] seeks to demonstrate economic benefits of using a probabilistic-dynamic approach (PDA) in the quantification of operating reserves as opposed to the traditional quantification methods. The PDA considers conventional generation outages; load and wind forecast uncertainty on an hourly basis; and load and wind variability in 10 min time frame. The authors objective was to determine the amount of operating reserves hourly in real time that minimizes the total cost of the power system, that is, the sum of operating costs and the socioeconomic costs related to the EENS.

In [20], authors address the critical problem of spinning reserve (SR) allocation for active power dispatch with large-scale wind power penetration. In particular the SR allocation for a multi-area power system where not well optimized SR allocation may make the inter-zonal reserve supplies constrained by the transmission interface limits jeopardizing the operational security of the power system. The objective of the study is to ensure operational risk is shared evenly among all sub-areas in SR allocation to avoid the possibility of a large blackout or wind power curtailment while considering economic and technical constraints.

A stochastic programming approach is used in [21] for quantifying reserve allocation based on scenarios generated by Monte Carlo Simulation. The objective was to demonstrate the benefits and drawbacks of the stochastic approach in reserve allocation by a comparison to deterministic approaches. This approach was found to be attractive as it enabled co-optimization of generation schedules and reserve requirement in the same problem formulation. It outperforms deterministic models in analyzing economic impacts of wind integration as it provides a means of assessing the sensitivity of operating costs in wind integration levels. However, it was found to present significant challenges in developing appropriate scenarios to be used as input to the stochastic programming model as well as computational intractability of the resulting problem.

In [22], the authors seek to address the problem of how to determine and schedule the additional SR requirement due to wind power fluctuation, more particularly how to jointly dispatch SR with power generation economically. Authors aim to investigate utilizing automatic generation control (AGC) units in both power supply to meet load demand as well as providing spinning reserve.

Using actual wind power and load data for two grids (ERCOT and MISO) in the US, authors in [23] aimed to demonstrate that both wind power and load forecast errors do not follow the normal distribution discrediting statistical approaches such as n-sigma criterion that assume this distribution in quantifying operating reserve. Authors demonstrate that using net load forecast error the dispatchable operating reserve requirement can be quantified based on certain reliability levels required.

B. Operational Constraints

In general operating reserve allocation problems consider the following constraints in their formulation.

LOLE threshold constraint: The resultant LOLE value after operating reserve allocation must be less than or equal to the incumbent LOLE to guarantee required levels of operational reliability of the power system.

Active power balance constraint: This requires that power balance between active power generation and load demand must be kept at all times during normal operations of the power system as expressed in the equation below

$$CG + Wind\ Power = Load\ Demand + Losses \quad (2)$$

Unit capacity constraint: Generation units usually have operational maximum and minimum power output limits within which the unit output must be maintained. *Ramp rate limits*: The ramp rate limits confine the power output increase or decrease between adjacent hours for certain units.

Transmission interface capacity constraints: The transmission capacity limits must be observed for accurate operating reserve allocation to ensure the reserve can be evacuated to demand centers without overloading the lines.

III. Solution Approaches

In the past several years, various researchers and organizations have participated or initiated wind power integration studies. The spine of each study is to evaluate the incremental need for additional operating reserves for the future system that result from high wind penetration. Basically study teams propose needed changes to maintain reliability while accommodating the variability and uncertainty present in the wind power. In most studies, statistical methods are used in the analysis and modelling of wind power time series data which forms a basis for quantification of operating reserve. Significant progress has been on methodologies used to compute theses values as successive studies addressed shortcomings of past studies. The most recent studies evaluating very high penetrations are using sophisticated methodologies that are diverging further from the traditional methods used today in actual system operations [3]. Generally methods to determine additional operating reserve can be classified as follows:

A. Statistical Approach

These rely on historical data of wind power and load which are analyzed to establish their statistical properties [7]. A commonly known statistical approach is the n-sigma criterion that is based on a comparison of the load and net load time series data, where net load is defined as load minus wind power production [9]. The probability distribution function of the net load is used to quantify additional operating reserve due to wind power integration at different time scales. Therefore assuming these data sets to be uncorrelated, the standard deviation of the net load time series is obtained as [4], [6], [8]

$$\sigma_{NL} = \sqrt{\sigma_L + \sigma_W} \quad (3)$$

where σ_L and σ_W are the standard deviations of the load and wind power time series respectively. From the analyzed time series of load and wind power data either variability or forecast error deviations can be established. The additional reserve ΔRes due to wind power integration can then be quantified according to [10]:

$$\Delta Res = n\left(\sigma_{NL} - \sigma_L\right) \quad (4)$$

The multiple n is defined a priori and represents the confidence level (CL) used. It varies from one power system to another. The remaining uncertainty not covered by n is met by energy balancing using the spot market.

According to literature review quite a number of studies have incorporated statistical approach in quantifying operating reserve requirement. In [24] authors proposed a model to incorporate the uncertainty of wind forecasts in unit commitments. Reference [6] presented a reserves procurement model that combines wind forecast uncertainty with load forecast uncertainty and generator forced outage probabilities.

Studies in [15], [16], each presented decision analytic frameworks to determine operating reserves in systems with wind power by balancing the cost of reserves and the cost of load curtailments. Authors in [25] presented an economic dispatch algorithm that accounts for wind forecast uncertainty. In these studies cited above statistical approach is applied, wind and load forecast errors are generally modeled using Gaussian distribution. On the contrary references [23], [26], [27] show that wind and load forecast errors are not normally distributed and demonstrate that wind forecast errors have distributions that can be highly skewed when the forecast is for a very low or high value of wind power. Wind forecasts near the middle of the forecast range have more symmetric error distributions (although not necessarily Gaussian). Fig 2 illustrates typical wind forecast errors as reported in [23]. The forecast horizon considered for wind power is therefore critical in determining more accurate results of operating reserve requirement using the statistical approach.

B. Probabilistic Approach

In more recent studies, reliability theory of the power system has been used to quantify operating reserve by integrating different sources of uncertainty [17]. Probabilistic approaches are based on system generation margin (SGM) which is defined as the difference between the total available generation and load demand. The SGM is a function of two random variable which makes it a random variable [15], [28]. In computation of SGM distribution, the probability distribution of CG, Wind Power generation W and load L must be considered.

$$SGM = CG + W - L \quad (5)$$

For a specific level of reserve R, the distribution of $SGM + R$ describes the probability that R is sufficient to cover the deficit in generation. Usually the SGM distribution forms a basis for calculation of classical reliability indices such as LOLE, LOLP and EENS [17], [29]. Operating reserve requirement are computed to meet these indices.

Authors in [18], present a two-level model that solves the allocation problem for composite power system. The upper model allocates operation reserve among the subsystems from

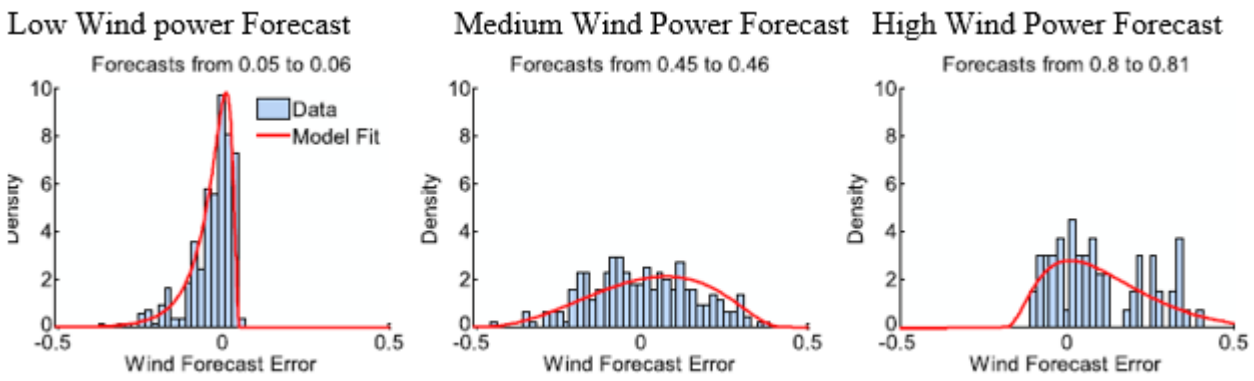


Fig. 2. Wind power forecast error distributions at three different wind forecasts [23].

the economic point of view while the lower model evaluates the system in the reserve schedule considering reliability levels. Optimization of the reserve allocation to achieve balance between reliability and economy is achieved through iteratively adjusting the reserve requirement by the upper model with the resulting reliability indices from the lower model.

The study team in [19] proposed a probabilistic-dynamic approach to quantify different operating reserve categories within a real-time system operation is based on an iterative process where the total costs (operating cost and cost of expected energy not supplied EENS) of the system are minimized. The PDA considers CG outages; load and wind forecast uncertainty on an hourly basis; and load and wind variability in 10 min time frame. The study demonstrates economic benefits of using a probabilistic-dynamic approach (PDA) in quantification of operating reserves as opposed to the traditional quantification methods in power systems with high penetration levels of wind power.

In [20], the authors propose a risk-based multi-objective SR optimization model which considers coordination of multiple control sub-areas to effectively handle the problem of local reserve inadequacy and transmission interface load. A probabilistic approach that considers forced outages of generators and forecast errors of load and wind power is used to establish relationship between SR allocation and LOLE. The resultant SR allocation problem is multi-objective. Fuzzy optimization method is used to transform the multi-objective optimization problem to a single-objective problem which can easily be solved. The single-objective problem characterized by nonlinearly coupled and not continuously differentiable constraints is solved using Particle Swarm Optimization (PSO) method.

A two-stage stochastic programming model for committing reserves in systems with large amounts of wind power is presented in [21]. The first stage formulates the UC problem of slow generators while the second stage formulates the hour-ahead economic dispatch problem for the entire system given the fixed day-ahead schedule of slow generators. Scenarios representing daily wind time series based on probability of occurrence are selected as part of input to the stochastic program. Lagrangian relaxation decomposition algorithm is used for solving the stochastic program.

Authors in [22] propose a joint optimization dispatch model of active power and additional SR requirement using AGC units for wind power integration systems. In formulation a model to accomplish this, both the probabilistic distributions of load and wind forecast errors were introduced to deal with uncertainties. Multiple scenarios are adapted to simulate the fluctuation of wind power and load. The actual output of AGC generators is adjusted according to the AGC unit modification strategy in each scenario.

Despite the attractive features presented by probabilistic approaches, the complexity of the formulated problem introduces computational intractability that can be rather time consuming to solve. However recent increased availability of high speed computing coupled with reduction in costs means that this problem can be overcome.

A summary of the general concept, advantages and disadvantages of the statistical approaches and probabilistic approaches are given in Table 1.

IV. Research Gaps

A review of literature on the sizing and allocation of additional operating reserve requirement due to large scale wind power integration as undertaken in this paper reveals that there has been commendable effort over the last decade or so in proposing solution methodologies and models to quantify and optimize allocation of operating reserve. Nevertheless, there exists a number of research gaps as outlined below:

1. An assumption of CL defined a priori still dominates the reviewed literature in quantifying operating reserve. Theres need for an objective method of determining this factor to guarantee consistency.
2. Most research done do not incorporate certain critical constraints such as network transmission capacity constraints in formulation of the reserve allocation problem. This undermines the practical applicability of the proposed methodologies.
3. Apart from ignoring important constraints, most work is silent on the different categories of operating reserve that can be deployed to meet the supply deficit as of wind power

TABLE I
A SUMMARY OF GENERAL CONCEPT, ADVANTAGES AND DISADVANTAGES OF SOLUTION APPROACHES

Approach	Concept	Advantages	Disadvantages
Probabilistic	• Based on system generation margin (SGM) - the difference between the total available generation and load demand. • Reliability theory of power systems is incorporated in quantifying of operating by integrating different sources of uncertainty. • Usually the SGM distribution forms a basis for calculation of classical reliability indices such as LOLE, LOLP and EENS and the operating reserve requirement are computed to meet these indices.	• Dynamic reserve allocation at desired time interval which eliminates the need to hold large amounts of reserve at every instant. • Allows incorporation of a variety of sources of uncertainty (variables) in determination of operating reserve which makes it more practical. • Uses an optimal CL determined based on a cost benefit analysis considering operating cost and reliability of the system.	• Computational intractability due to the complexity of the resultant problem. • Significant challenge presented in selecting appropriate scenarios and their probabilities.
Statistical	• Relies on historical data of wind power and load which are analyzed to establish their statistical properties. • A commonly known statistical approach is the n-sigma criterion where the probability distribution function of the net load (Load − Wind Power) is used to quantify additional operating reserve due to wind power integration at different time scales.	• Results in relatively simple problem formulation hence provides quick solutions to operating reserve determination. • Useful as an easier way of providing an approximation close to reality of reserve determination.	• Static reserve allocation which results in unnecessarily holding of large amounts of reserve thereby increasing costs. • Assumes WFE and variability follows Gaussian distribution which is not necessarily the case. • It is implicitly assumed that present and future behavior of the power system resembles observations of the past. If this assumption is not valid, the n-sigma criterion or any statistical approaches will fail.

fluctuation which is equally essential in minimizing cost and avoiding wind power curtailment.

V. CONCLUSION

This paper presents an overview of the methodologies used in the sizing and allocation of additional operating reserve requirement with large scale wind power integration. Recent reviewed literature on this subject reveals that significant progress has been made in the quantification of additional reserve. Generally the solution methods to the reserve problem formulated are based on statistical or probabilistic approaches. However, there are still some inconsistencies with data and methodologies used where improvements can be made. Finally, a number of research gaps have been identified such as the need to incorporate critical system constraints such as network transmission capacity constraints in the problem formulation; limited direction on the different categories of reserves to be deployed after quantification; and lack of consensus on the CL to be used on quantification.

REFERENCES

[1] X. Liu and W. Xu, "Minimum emission dispatch constrained by stochastic wind power availability and cost," *IEEE Transactions on Power Systems*, vol. 25, pp. 1705–1713, Aug 2010.

[2] D. J. Maggio, "Impacts of wind-powered generation resource integration on prices in the ercot nodal market," in *2012 IEEE Power and Energy Society General Meeting*, pp. 1–4, July 2012.

[3] E. Ela, B. Kirby, E. Lannoye, M. Milligan, D. Flynn, B. Zavadil, and M. O'Malley, "Evolution of operating reserve determination in wind power integration studies," in *IEEE PES General Meeting*, pp. 1–8, July 2010.

[4] L. Wang, A. Gerber, J. Liang, L. Dong, and X. Liao, "Wind power forecasting for reduction of system reserve," in *45th International Universities Power Engineering Conference UPEC2010*, pp. 1–5, Aug 2010.

[5] R. Piwko, X. Bai, K. Clark, G. Jordan, N. Miller, and J. Zimberlin, "The effects of integrating wind power on transmission system planning, reliability, and operations."

[6] R. Doherty and M. O'Malley, "A new approach to quantify reserve demand in systems with significant installed wind capacity," *IEEE Transactions on Power Systems*, vol. 20, pp. 587–595, May 2005.

[7] C. Hamon and L. Sder, "Review paper on wind power impact on operation of reserves," in *2011 8th International Conference on the European Energy Market (EEM)*, pp. 895–903, May 2011.

[8] R. Doherty and M. O'Malley, "Quantifying reserve demands due to increasing wind power penetration," in *2003 IEEE Bologna Power Tech Conference Proceedings,*, vol. 2, pp. 5 pp. Vol.2–, June 2003.

[9] H. Holttinen, "Impact of hourly wind power variations on the system operation in the nordic countries," *Wind Energy*, vol. 8, no. 2, pp. 197–218, 2005.

[10] H. Holttinen, M. Milligan, B. Kirby, T. Acker, V. Neimane, and T. Molinski, "Using standard deviation as a measure of increased operational reserve requirement for wind power," *Wind Engineering*, vol. 32, no. 4, pp. 355–378, 2008.

[11] P. Srensen and N. A. Cutululis, "Wind farms' spatial distribution effect on power system reserves requirements," in *2010 IEEE International Symposium on Industrial Electronics*, pp. 2505–2510, July 2010.

[12] P. Norgaard and H. Holttinen, "A multi-turbine power curve approach," in *Nordic Wind Power Conference, Gothenburg*, pp. 1–5, 1-2 March 2004.

[13] G. Strbac, A. Shakoor, M. Black, D. Pudjianto, and T. Bopp, "Impact of wind generation on the operation and development of the uk electricity systems," *Electric Power Systems Research*, vol. 77, no. 9, pp. 1214 – 1227, 2007. Distributed Generation.

[14] G. Liu and K. Tomsovic, "Quantifying spinning reserve in systems with significant wind power penetration," *IEEE Transactions on Power Systems*, vol. 27, pp. 2385–2393, Nov 2012.

[15] M. A. Matos and R. J. Bessa, "Setting the operating reserve using prob-

abilistic wind power forecasts," *IEEE Transactions on Power Systems*, vol. 26, pp. 594–603, May 2011.

[16] M. A. Ortega-Vazquez and D. S. Kirschen, "Estimating the spinning reserve requirements in systems with significant wind power generation penetration," in *2009 IEEE Power Energy Society General Meeting*, pp. 1–1, July 2009.

[17] R. Billinton, R. N. Allan, and R. N. Allan, *Reliability Evaluation of Power Systems*. Plenum Press, 2 ed.

[18] J. Wang, A. F. Zobaa, C. Huang, and C. Chen, "Day-ahead allocation of operation reserve in composite power systems with large-scale centralized wind farms," *Journal of Modern Power Systems and Clean Energy*, vol. 4, pp. 238–247, Apr 2016.

[19] C. Rahmann, A. Heinemann, and R. Torres, "Quantifying operating reserves with wind power: towards probabilistic dynamic approaches," *IET Generation, Transmission Distribution*, vol. 10, no. 2, pp. 366–373, 2016.

[20] J. Chen, W. Wu, B. Zhang, B. Wang, and Q. Guo, "A spinning reserve allocation method for power generation dispatch accommodating large-scale wind power integration," vol. 6, pp. 5357–5381, 10 2013.

[21] A. Papavasiliou, S. S. Oren, and R. P. O'Neill, "Reserve requirements for wind power integration: A scenario-based stochastic programming framework," *IEEE Transactions on Power Systems*, vol. 26, pp. 2197–2206, Nov 2011.

[22] S. Xia, M. Zhou, G. Li, Y. Liu, and M. Xiang, "On spinning reserve determination and power generation dispatch optimization for wind power integration systems," in *2012 IEEE Power and Energy Society General Meeting*, pp. 1–7, July 2012.

[23] B. Mauch, J. Apt, P. M. S. Carvalho, and P. Jaramillo, "What day-ahead reserves are needed in electric grids with high levels of wind power?," *Environmental Research Letters*, vol. 8, no. 3, p. 034013, 2013.

[24] J. Wang, M. Shahidehpour, and Z. Li, "Security-constrained unit commitment with volatile wind power generation," in *2009 IEEE Power Energy Society General Meeting*, pp. 1–1, July 2009.

[25] F. Bouffard and F. D. Galiana, "Stochastic security for operations planning with significant wind power generation," in *2008 IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, pp. 1–11, July 2008.

[26] H. Bludszuweit, J. A. Dominguez-Navarro, and A. Llombart, "Statistical analysis of wind power forecast error," *IEEE Transactions on Power Systems*, vol. 23, pp. 983–991, Aug 2008.

[27] M. Lange, "On the uncertainty of wind power predictionsanalysis of the forecast accuracy and statistical distribution of errors," vol. 127, 05 2005.

[28] M. A. Matos and R. Bessa, "Operating reserve adequacy evaluation using uncertainties of wind power forecast," in *2009 IEEE Bucharest PowerTech*, pp. 1–8, June 2009.

[29] R. J. Bessa and M. A. Matos, "Comparison of probabilistic and deterministic approaches for setting operating reserve in systems with high penetration of wind power," in *7th Mediterranean Conference and Exhibition on Power Generation, Transmission, Distribution and Energy Conversion (MedPower 2010)*, pp. 1–9, Nov 2010.

Overview of Congestion Management in Deregulated Power System: Review

Mahouna Houndjéga, Christopher M. Maina and Cyrus W. Wekesa

Abstract— The rise in competition in the energy market has enabled an optimal transmission system in the electricity sector. The utilization of the transmission system makes some lines congested due to some constraints capacities of the line, which is a big problem for the system operator. Congestion Management is one of the main issues that threaten the system security and the most challenging tasks System Operator in deregulated Power System. A lot of attention has been given to the congestion management in the last years due to its benefits such as having a secure and reliable system. Congestion is the situation when technical constraints or economic restrictions are violated, which has a negative effect on the electricity market. The congestion management permits to alleviate the violated constraint. System operators try to manage congestion, which otherwise increases the cost of the electricity and also threatens the system security and stability. Several methods are proposed to solve the issue of congestion. This paper presents the overview and comparative studies of the research methods that are used for Congestion Management in deregulated power systems.

Keywords— Congestion management, Electricity market, System operator, transmission system**.**

I. INTRODUCTION

With the growth of the population in the world and the increased of the electricity demand, the transmission system plays an important role however the existence of some constraints of it. There is a quantity of power that can be flowed between two buses at different points of the grid. This does not permit to deliver all the amount of power for the different contracts, bilateral, multilateral and to fulfil the pool demand. Forcing to satisfy all the demand leads the violation of some constraints such as voltage limit and line limit. It is referred to as congestion. Congestion whether in monopoly system or in deregulated system cannot be allowed. It leads the system not be secure and reliable. Then, congestion management has become the center of many researches in deregulated electricity market. According to reference [1], congestion in transmission systems is a situation where the demand for transmission capacity exceeds the transmission grid capabilities; this condition might result in a violation of network security limits, such as thermal, voltage stability, or angular stability.

Mahouna H., *Pan African University, Institute for Basic Sciences, Technology and Innovations (Pauisti/Jkuat).* (phone: +254789757491; e-mail: dj.mahouna@gmail.com).

II. PROBLEM FORMULATION AND CONSTRAINTS

Congestion in deregulated power systems is an optimization problem. The most commonly method used is the minimization of the total cost of congestion. Congestion management with total cost of congestion minimization can be evaluated by [2][3][4]

Minimize $TC = \sum_{j=1}^{ng}(C_k\,\Delta Pgj^{+} + D_k \Delta Pgj^{-})$ $\$/hr$ (1)

TC: total congestion cost in $/hr,
C_k : Incremental biding cost,
D_k : Decremental biding cost,
ΔPgj^{+} : Amount of active power increment in the generator j
ΔPgj^{-} : amount of active power decrement in the generator j.

The most commonly constraints used are shown in [3]
Equality constraints:
Power Balance

$P_{gi} - P_{di} - \sum_{j=1}^{N}|V_i|\,|V_j||Y_{ji}|\cos(\delta_i - \delta_j - \theta_{ij}) = 0$ (2)

$Q_{gi} - Q_{di} - \sum_{j=1}^{N}|V_i|\,|V_j||Y_{ji}|\sin(\delta_i - \delta_j - \theta_{ij}) = 0$ (3)

$P_{gi} = P_{gi}^{c} + \Delta Pgj^{+} - \Delta Pgj^{-}$;i=1,2,…,ng (4)

$P_{dk} = P_{dk}^{c} \quad ; k = 1,2,\ldots\ldots, Nd$ (5)

Inequality constraints:
Power generation limit:

$P_{Gi}^{min} \le P_{Gi} \le P_{Gi}^{max}$,i=1,…..,ng (6)

$Q_{Gi}^{min} \le Q_{Gi} \le Q_{Gi}^{max}$, i=1,…..,ng (7)

Incremented or decremented real power limit:

$(P_{gi} - P_{Gi}^{min}) = \Delta P_{Gi}^{min} \le \Delta P_{Gi} \le \Delta P_{Gi}^{max} = (P_{Gi}^{max} - P_{gi})$ (8)

$\Delta Pgj^{+} \ge 0; \quad \Delta Pgj^{-} \ge 0$ (9)

III. CONGESTION MANAGEMENT METHODS

Many methods have been used in congestion management. Two of them are fluently used:

A- Flexible AC Transmission System (FACTS) Device

Facts device is used in congestion management for many purposes. Reference [5] proposed a multi-objective for congestion management by locating and sizing Series FACTS and the newly developed Grey Wolf Optimizer. Optimal location of TCSC was proposed in [6] using PSO. In [7], the authors proposed a review article on the various publications on congestion management in past few decades using FACTS

Muriithi C M , School of Engineering and Technology, Murang'a University of Technology (e-mail: cmmuriithi@mut.ac.ke).
Wekesa C, School of Electrical & Electronic Engineering, Technical University Of Kenya (email: cyrus.wekesa@gmail.com)

devices and the Study of Various Cases for Congestion Management in Different Countries. In [8], it performed the best location of FACTS device like TCSC and UPFC by using sensitivity based Eigen value analysis and the performance analysis. The proposed approach had a capability to enhance the Voltage stability, small signal stability and minimize the real power loss in the power systems network. The authors presented an effective approach for optimal allocation and optimal sizing of Thyristor Controlled Series Compensator (TCSC) to reduce the congestion in day-ahead power market. Active Power Spot Price Index (APSPI) was used to reduce the solution space effectively and to determine the best location of TCSC in the system[9]. Reference [10] proposed also a review of congestion management by deciding optimal location of Facts device. They proposed new indices to voltage drop compensation, also congestion rent contribution method. Comparative studies in restructured power system has been proposed in [11] by the optimal placement of TCSC using sensitivity approach and pricing method.

B- Optimization techniques and expert system

Congestion problem is non-linear problem. Many optimization techniques and expert systems have been used in congestion management.

In [12] Genetic Algorithm based rescheduling of generators is developed to alleviate the congestion. In reference [13],the authors proposed an efficient method for transmission line over load alleviation in deregulated power system. The sensitivity factor of the congested line is used to select the generators and the active power of the participating generators is rescheduled using Cuckoo Search (CS) algorithm for relieving congestion. In the other hand, [14] presented a Genetic Algorithm (GA) based new reconfiguration algorithm of the network which will able to identify the most congested area of power network and fabricate the least loss condition after alleviating overload and overvoltage as well as ensuring efficient network operation.

According to [15], the authors used adaptive real coded biogeography based optimization to minimize rescheduling power and hence minimize the congestion cost. In the other hand, Intelligent method is proposed for congestion management in power system in [16]. According to reference [17], congestion management in hybrid electricity market for hydro-thermal was proposed. In [18]-[4] , PSO algorithm was used to minimize the reschedule of the generator Output with different objectives functions. A real coded genetic algorithm was used to find the optimal generation rescheduling for relieving congestion [19]. According to reference [20], the authors proposed a new model for power system CM by considering power system uncertainties based on chance-constrained . For the reference [21], congestion relief procedure has been discussed and compared with the objective of rescheduling cost minimization and proposed objective of real power loss minimization. Modified Grey Wolf Optimizer Used for Congestion Management in a Deregulated Power Systems was proposed [22]. The authors [23] proposed an approach considering the Risk of Cascading Failures for congestion management. In the other hand, reference [24] proposed a novel algorithm called DEKH algorithm in the congestion management in a deregulated system with market analysis. It proposed that the market analysis calculates the impact factors in each zone. The major contribution is, it utilizes both active and reactive power cost functions in the objective function since the reactive power plays a vital role in the congestion relief at low congestion cost. In the other hand, reference [25] proposed a novel PSO strategy for transmission congestion management. In this research, Particle Swarm Optimization (PSO) with varying inertia weight strategy, with two variants e1-PSO and e-2 PSO are applied for optimal solution of active and reactive power rescheduling for managing congestion. The generators sensitivity technique is opted for identifying participating generators for managing congestion.

In the other hand in [26] , the authors proposed a probabilistic model to reduce the probability of line congestions and voltage violations in a smart grid located in a radial distribution network. Renewable distributed resources and a high penetration of Electric Vehicles (EVs) are considered. A probabilistic power flow based on the point estimate method is firstly used to compute the distribution functions of the line flows and node voltages. Next, a congestion management strategy is proposed in order to keep the line flows and node voltages within the appropriate range at a given confidence level. The strategy is based on the sensitivity distribution factors: Power Transfer Distribution Factors (PTDFs) and Voltage Distribution Factors (VDFs). Reference [27] proposed a flower pollination algorithm (FPA) for congestion management (CM) problem of deregulated electricity market. Nature-inspired FPA is based on the characteristics of pollination process of flowering plants. The aim of employing FPA is to effectively relieve congestion in the transmission line by means of rescheduling of real power output of the generators. In the other hand a new method to manage congestion based on low power tracing was proposed in reference [28]. In power systems, lines and transformers transmit energy which is a comprehensive reflection of the active power and reactive power. Thus the causes of congestion should include the active and reactive power. The results of sensitivity analysis method may be inaccurate because only the effect of active power is considered. When there exists congestion, the analytical relationships between the line power and generator outputs can be obtained by power flow tracing. Then the power vectorgraph is made, and congestion can be eliminated effectively by adjusting relevant power components in four quadrants. Penguin search optimization algorithm is used by Ravish Yadav to reduce congestion on IEEE-30 bus system & its comparison to PSO (Particle swarm optimization). Although PSO also reduces congestion but PeSOA is better than PSO in the reference [29].

In the other hand, an obvious technique of congestion management is rescheduling the power outputs of generators in the system. Generation sensitivity factor has been used to identify the generators, which affects more on the congested line. However, all generators in the system need not to take part in congestion management [30].

In reference [31], a contribution has been made with another technique for reliving the congested power flow in a transmission line has been introduced using Fuzzy Logic with Interline Power Flow Controller. For this approach a fuzzy logic controller are proposed to control the interline power flow controller device depending on real and reactive

power. Fuzzy logic controller is uncertainty, which has been developed to solve the congestion management problem

In [32], the authors used the locational dependence of congestion cost and a comparison of the voltage stability margin based approach. For them, system security needs to be ensured especially in a hybrid network. Renewable Energy Sources are present which might disturb the network due to their intermittent nature. Voltage stability margin based congestion management strategy based on scheduling of generator and demand unit as per their effective bids. In the other hand, Distributed Generation (DG) is used on congestion management in Deregulated Electricity Market. Two different methods for optimal placement of DG to manage congestion are proposed which are based on the highest locational marginal price of buses and difference of locational marginal prices between two buses in reference [33]. According to [34] the authors presented the various methods of congestion reduction in Indian power sector. The authors proposed in reference [35] a method using optimal power flow topology for congestion management in power system. In reference [36] the authors developed a method using market splitting based approach for relieving congestion. In reference [37] the authors employed an Heuristic search algorithms incorporating wireless technology method to minimize the congestion cost. In reference [38], an approach using Generation rescheduling to relieve congestion is proposed. It is formulated as an optimal power flow (OPF) and solved by employing Particle Swarm Optimization.

The authors in reference [39] tried to find the optimal rescheduling of active power generations based on real power sensitivity index of the generators so as to minimize the congestion cost. In [40], the authors proposed the using of a novel Satin bowerbird optimization for real power rescheduling of generators for congestion management.

In reference [41], the authors proposed the application of MATPOWER for the analysis of congestion and its using to determine the generator sensitivity factor.

The authors developed in reference [2] an optimal real power rescheduling of generation using a novel ant lion optimizer for congestion. Changing the pattern of real power generation from generators are used for congestion management and black hole algorithm (BHA) is used for identifying the optimal generation pattern for avoiding congestion in reference [3]. In reference [42], real and reactive power rescheduling based congestion management is used to relieve the transmission congestion. To have an optimal rescheduling of the generator, black hole algorithm (BHA) is performed.

In reference [43] a Combined Economic and Emission Dispatch (CEED) by employing a novel technique of optimization through Artificial Bee Colony (ABC) algorithm. In reference [44] generator rescheduling is proposed as the congestion management technique. The authors used the generator sensitivity factor to identify the generator participating in congestion management and Firefly algorithm is used to find the optimal rescheduling. An Improved Differential Evolution (IDE) algorithm was presented in reference [45] to alleviate Congestion in transmission line by rescheduling of generators while considering voltage stability. In reference [46] Artificial Bee Colony algorithm was used for real power rescheduling to relieve congestion. The authors computed the generator sensitivity factor for the congested lines.

The authors proposed in reference [47] a concept of transmission congestion penalty factors and its implementation to control power overflows in transmission lines for congestion management. They used a Re-dispatch methodology for cost of transmissions network to its user.

In reference [48] the congestion management of Locational Marginal Pricing (LMP) with minimum line loss, less cost, line flow, better sensitivity and better performances in optimal power flow and control flow. In reference [49] it deals with the power trading in electricity market to ensure regular supply at competitive rates with multi congestion case. They proposed a method to compute MCP and LMP for Pool based transaction.

Reference [50] made a review on congestion management in the deregulated power market. It reviews some of the congestion management (CM) methods including the nodal pricing method, differential evolution (DE), addition of renewable energy sources, extended quadratic interior point (EQIP) based OPF, mixed integer nonlinear programming, particle swarm optimization (PSO), cost-free methods and Genetic Algorithm (GA). In [51] an overview of transmission congestion occurring in the deregulated electricity market is presented. In this paper, are presented the definitions and the reasons due to which transmission congestion occurs. Reference [52] presented a review on congestion and its effects on deregulated power system market. They made a presentation of the methods used in congestion management.

The proposed paper, Varaprasad Janamala developed some of the technical and non-technical approaches using simple case studies[53].

IV. RESEARCH NEEDS

This paper presents a review of the literature based on congestion management in deregulated environment. Many efforts have been done over the last decade in research to try various solution methodologies for solving congestion which is a very complex optimization problem. Nonetheless, a number a research needs exist as outlined below:

1- Most research considers only the minimization of the congestion cost and ignore that there is no cost for congestion in deregulated electricity market. It is important to forecast the congestion then the system operator will not have to put any cost for congestion management.

2- With the development of novel optimization algorithm, this will permit to have a good performance when we have the congestion in the system.

V. CONCLUSION

Congestion management became a center of interest in deregulated electricity market. It permits to secure the grid from the violation of some constraints. This paper presents a summary of some previous works in the area of congestion management. It constitutes a tool for future work in the same area. From the previous works, the main method used in congestion management is the rescheduling of the generators output with minimum cost congestion. For many works, IEEE 30 bus system test is used for different case studies. Congestion

can be due to the outage of line, the sudden increased of the load or again the reduction of the line limit.

VI. Appendix

Comparison of different optimization algorithms for Congestion Management applied on IEEE 30 Bus System.

TABLE I
SIMULATION RESULTS FOR DIFFERENT ALGORITHMS FOR IEEE 30-BUS SYSTEM

No	Congestion Management Methods	Type of contingency	CM Cost ($/MWH)	Referenc
1	Particle Swarm Optimization	Outage of line 1-2	538.95	[4]
2	Novel Satin Bowerbird		421.58	[40]
3	Novel flower pollination		519.62	[27]
4	Artificial Bee Colony Algorithm		2867.3	[46]
5	Improved Differential Evolution		1655.65	[45]

The IEEE 30 Bus system data is in [4].

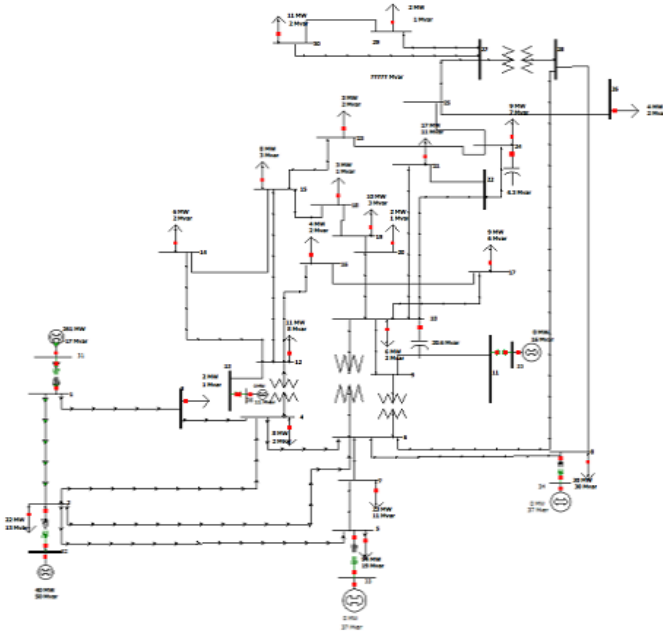


Fig. 1: IEEE 30-bus system

Acknowledgement

This research is funded by Pan African University under African Union research grant.

References

[1] H. Khani, S. Member, M. R. D. Zadeh, S. Member, A. H. Hajimiragha, and S. Member, "Owned Large-Scale Energy Storage Systems in a," *IEEE Trans. POWER Syst. Charg.*, pp. 1–10, 2015.

[2] S. Verma and V. Mukherjee, "Optimal real power rescheduling of generators for congestion management using a novel ant lion optimiser," *IET Gener. Transm. Distrib.*, vol. 10, pp. 2548–2561, 2016.

[3] R. R and A. M, "Sensitivity Based Optimal Real Power Rescheduling for Management Using Black Hole Algorithm," *Aust. J. Basic Appl. Sci.*, vol. 10, no. October, pp. 183–193, 2016.

[4] N. K. Sujatha Balaraman, "Transmission Congestion Management Using Particle Swarm Optimization," *Journal of Electrical Systems*, vol. 7, no. 1. pp. 54–70, 2011.

[5] A. R. Moradi, Y. Alinejad-Beromi, and K. Kiani, "Locating of series FACTS devices for multi-objective congestion management using components of nodal prices," *Iran. J. Electr. Electron. Eng.*, vol. 13, no. 1, pp. 32–46, 2017.

[6] M. Mandala and C. P. Gupta, "Transmission congestion management with TCSC using Bacterial Foraging-Particle Swarm Optimization," *PEDES 2012 - IEEE Int. Conf. Power Electron. Drives Energy Syst.*, pp. 1–6, 2012.

[7] M. Gupta, V. Kumar, G. K. Banerjee, and N. K. Sharma, "Mitigating Congestion in a Power System and Role of FACTS Devices," *Adv. Electr. Eng.*, vol. 2017, pp. 1–7, 2017.

[8] R. Retnamony and I. J. Raglend, "N-1 Contingency analysis in a Congested power system and enhance the voltage stability using FACTS Devices," *I J C T A*, vol. 9, no. 7, pp. 3147–3157, 2016.

[9] S. Dawn, "An Active Power Spot Price based Approach for Congestion Management by Optimal Allocation of TCSC in Competitive Power Market Za," *IEEE Reg. 10 Conf. -Proceedings Int. Conf.*, pp. 1480–1484, 2016.

[10] D. Kharate, R. K. Verma, and S. B. Mohod, "A review on Congestion Management of Transmission Line by Deciding Optimal Location of Facts Device," no. 5, pp. 1988–1992, 2017.

[11] J. S. Rao and J. Amarnath, "Transmission Congestion Management Comparative Studies In Restructured Power System," *Int. J. Sci. Eng. Res.*, vol. 4, no. 8, pp. 177–182, 2013.

[12] R. Harish and G. Kannan, "Congestion Management by Generator Rescheduling using Genetic Algorithm Optimization Technique," *Int. J. Adv. Res. Electr. Electron. Instrum. Eng.*, no. 4, pp. 50–58, 2014.

[13] N. Chidambararaj and K. Chitra, "Congestion Management Based On Active Power Rescheduling Of Generator units using Cuckoo Search Algorithm," *Int. J. Sci. Eng. Res.*, vol. 5, no. 4, pp. 33–38, 2014.

[14] S. Pal, S. Sen, and S. Sengupta, "Power Network Reconfiguration For Congestion Management And Loss Minimization," *Michael Faraday IET Int. Summit MFIIS-2015*, 2015.

[15] A. Ramesh Kumar and L. Premalatha, "Congestion management in open access electric network using adaptive real coded biogeography-based optimization," *ARPN J. Eng. Appl. Sci.*, vol. 10, no. 12, pp. 5213–5222, 2015.

[16] S. A. H. Zadeh, A. A. Gharaveisi, and G. H. Shahgholian, "Congestion Management in Power Systems Via Intelligent Method," *Int. J. Smart Electr. Eng.*, vol. 2, no. 1, pp. 23–31, 2013.

[17] Y. P. Verma and A. Kumar, "Congestion Management in Hybrid Electricity Market for Hydro-Thermal System," in *Proceedings of the World Congress on Engineering 2014*, 2014, vol. I.

[18] S. Dutta and S. P. Singh, "Optimal Rescheduling of Generators for Congestion Management Based on Particle Swarm Optimization," *IEEE Trans. Power Syst.*, vol. 23, no. 4, pp. 1560–1569, 2008.

[19] S. Balaraman, "Congestion Management in Deregulated Power System Using Real Coded Genetic," *Int. J. Eng. Sci. Technol.*, vol. 2, no. 11, pp. 6681–6690, 2010.

[20] M. Hojjat, M. H. Javidi, and D. Yazdanpanah, "Stochastic congestion management considering power system uncertainties: A chance-constrained programming approach," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 24, no. 1, pp. 61–75, 2016.

[21] S. Charles Raja, P. Venkatesh, and B. V. Manikandan, "Transmission Congestion Management in restructured power systems," *2011 Int. Conf. Emerg. Trends Electr. Comput. Technol. ICETECT 2011*, pp. 23–28, 2011.

[22] R. Retnamony, J. Raglend, K. Kumari, T. Nadu, and T. Nadu, "Modified Grey Wolf Optimizer Used for Congestion Management

in a Deregulated Power Systems," vol. 116, no. 22, pp. 315–326, 2017.

[23] J. Hazra and D. P. Seetharam, "A network congestion management approach considering the risk of cascading failures," *2010 Jt. Int. Conf. Power Electron. Drives Energy Syst. PEDES 2010 2010 Power India*, 2010.

[24] M. minnet;Faustina Chidambararaj, "Congestion Management in a Deregulated System with Market Analysis Using DEKH Algorithm," *Int. J. Tech. Res. Appl.*, vol. 4, no. 5, pp. 105–110, 2016.

[25] H. Mahala and Y. Kumar, "Novel Pso Strategy For Transmission Congestion Management," *Electr. Electron. Eng. An Int. J.*, vol. 5, no. 2, pp. 1–9, 2016.

[26] S. Martin, J. A. Aguado, and S. De Torre, "Probabilistic congestion management using EVs in a smart grid with intermittent renewable generation," *Electr. Power Syst. Res.*, vol. 137, pp. 155–162, 2016.

[27] S. Verma and V. Mukherjee, "A novel flower pollination algorithm for congestion management in electricity market," *3rd Int'l Conf. Recent Adv. Inf. Technol. | RAIT-2016*, pp. 1–6, 2016.

[28] H. Wang and H. Bao, "A New Method of Congestion Management Based on Power Flow Tracing," *China Int. Conf. Electr. Distrib. (CICED 2016)*, no. Ciced, pp. 10–13, 2016.

[29] R. Yadav, "Congestion Management in Dergulated Market along with the Promotion of Renewable Energy Resources Using PeSOA Optimisation," *Int. J. Sci. Res.*, vol. 6, no. 8, pp. 862–867, 2017.

[30] D. Choudhary and N. Kumar, "Congestion Management in Deregulated Power by Rescheduling of Generators," *Int. Res. J. Eng. Technol.*, pp. 1690–1695, 2017.

[31] P. Nirmalraj and A. Amsavalli, "Congestion Management in a Deregulated Power System Using Fuzzy Logic with IPFC," vol. 3, no. Iv, pp. 3–8, 2015.

[32] R. Sagwal and A. Kumar, "Congestion Management Solution for Hybrid System considering Voltage Stability Margin," *Procedia Technol.*, vol. 25, no. Raerest, pp. 726–734, 2016.

[33] A. S. Siddiqui, "Congestion Management in Deregulated Electricity Market Using Distributed Generation," no. 1, pp. 3–7, 2015.

[34] Q. Khan, F. Ahmad, and M. Imran, "Congestion management in Indian Power Transmission System," *Int. J. Eng. Technol.*, vol. 9, no. 3, pp. 26–31, 2017.

[35] G. M. Kumar, P. V Satyaramesh, K. S. R. Anjaneyulu, P. Sujatha, and A. Pradesh, "Congestion Management In Power System Using Optimal Power Flow Topology," *Int. J. Electr. Electron. Eng.*, vol. 6, no. 1, pp. 1–14, 2017.

[36] M. Harchand and KanwardeepSingh, "Congestion Management in Deregulated Power System Using Market Splitting Based Approach," *Int. J. Eng. Innov. Technol.*, vol. 7, no. 1, pp. 42–48, 2017.

[37] G. S. Jasmine and P. Vijayakumar, "Congestion Management in Deregulated Power System Using Heuristic Search Algorithms Incorporating Wireless Technology," *Wirel. Pers. Commun.*, vol. 94, no. 4, pp. 2665–2680, 2017.

[38] S. Sivasankari and N. Chidambararaj, "Generation Rescheduling Based Congestion Management in Restructured Power System," *Middle-East J. Sci. Res.*, vol. 25, no. 4, pp. 859–863, 2017.

[39] O. A. Journal and M. Arun, "Optimal Active Power Rescheduling of Generators for Congestion Management Based on Sensitivity Factors," vol. 10, no. 15, pp. 13–22, 2016.

[40] J. Chintam and M. Daniel, "Real-Power Rescheduling of Generators for Congestion Management Using a Novel Satin Bowerbird Optimization Algorithm," *Energies*, vol. 11, no. 1, p. 183, 2018.

[41] K. Paul and N. Kumar, "Application of matpower for the analysis of congestion in power system network and determination of generator sensitivity factor," *Int. J. Appl. Eng. Res.*, vol. 12, no. 6, pp. 969–975, 2017.

[42] R. Ramachandran and M. Arun, "Optimal Rescheduling of Real and Reactive Power for Congestion Management under Competitive Environment," *Int. J. Adv. Eng. Res. Sci.*, vol. 6495, no. 10, 2016.

[43] A. A. Christy, K. Balamurugan, P. A. D. V. Raj, and S. Rajasomashekar, "A Novel Optimization Technique for Combined Economic and Emission Dispatch Problem considering Congestion Management," *Indian J. Sci. Technol.*, vol. 9, no. October, 2016.

[44] S. Gope and S. Deb, "Generator Rescheduling for Congestion Management using Firefly Algorithm," *Int. Conf. Energy Syst. Appl.*, no. Icesa, pp. 40–44, 2015.

[45] S. T. Suganthi, D. Devaraj, and S. H. Thilagar, "An improved differential evolution algorithm for congestion management considering voltage stability," *Indian J. Sci. Technol.*, vol. 8, no. 24, 2015.

[46] R. Saranya, K. Balamurugan, and M. Karuppasamypandiyan, "Artificial Bee Colony Algorithm Based Congestion Management in Restructured Power System," *Indian J. Sci. Technol.*, vol. 8, no. April, pp. 171–178, 2015.

[47] A. Jaisisngpure, V. K. Chandrakar, and R. M. Mohril, "Market Based Criteria for Congestion Management and Transmission Pricing," vol. 4, no. 7, pp. 18–23, 2014.

[48] K. B. R. Kumar and S. C. Mohan, "LMP Congestion Management Using Enhanced STF-LODF in Deregulated Power System," *Sci. Res. Publ.*, pp. 2489–2498, 2016.

[49] S. K. Gupta, R.Bansal, Partibha, and M. Saini, "Review of Energy Technologies and Policy Research ENERGY TRADING AND CONGESTION MANAGEMENT USING REAL AND REACTIVE POWER RESCHEDULING AND LOAD CURTAILMENT Contribution / Originality," *Rev. Energy Technol. Policy Res.*, vol. 1, no. 2, pp. 28–41, 2014.

[50] S. Sharma, M. Kashyap, and SatishKansal, "Congestion Management in Deregulated Power Market – a Review," *Int. Conf. Adv. Emerg. Technol.*, no. Icaet, pp. 18–23, 2016.

[51] D. Kharate, R. K. Verma, and S. B. Mohod, "A review on Congestion Management of Transmission Line by Deciding Optimal Location of Facts Device," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 3, no. 5, pp. 1988–1992, 2017.

[52] A. K. Loveleen Kaur, Aashish Ranjan, S.Chatterji, "Congestion management approaches in deregulated power system – an illustrative approach," *Int. J. Grid Distrib. Comput.*, vol. 9, no. 12, pp. 171–180, 2016.

[53] V. Janamala, "Congestion Management Approaches in Deregulated Power System – An Illustrative Approach," *Int. J. Grid Distrib. Comput.*, vol. 9, no. 12, pp. 171–180, 2016.

Classification of drying methods for macadamia nuts based on the glcm texture parameters

Simon N. Njuguna[1,2], Stephen Ondimu[2], Glaston M. Kenji[3]

Abstract**—**Texture is an important feature in the definition of the external appearance of a material. In this study, Gray-level co-occurrence matrix (GLCM) was used in the analysis of image texture without the destruction of the two varieties of macadamia nuts , KRG-15 and MRG-20, that are dried using seven different methods. Using the Imagej software, four GLCM features were computed in order to define the effect on drying methods on quality of the two varieties. These GLCM features were contrast, Angular Second Moment (ASM), homogeneity and correlation. The aim of this study was to investigate the influence of drying methods on the quality of macadamia nuts for the two varieties: - KRG-15 and MRG-20 with reference to texture quality. The highest classification accuracy of 77% was attained for correctly identifying influence of drying method at direction at angle of 0˚ and displacement, d=1. Solar tent drying method produces superior quality macadamia nuts for both varieties comparable to drying methods; with GLCM feature values (contrast, Homogeneity, Correlation and ASM) of 8,179, 0.048, 0.798 and 0.009 respectively for KRG-15 and 10.238, 0.789, 0.007 and 0.047 respectively for MRG-20. Combination drying of solar tent and Microwave drying produce the least quality,

Keywords—Image texture, GLCM, Angular Second Moment, Homogeneity

1. INTRODUCTION

Macadamia nuts grown in Kenya originated from Australia. Other countries growing macadamia nut are U.S.A, which is the leading world producer of the nuts, followed by Australia, Kenya and South Africa [1].

The annual production of these nuts in Kenya are approximated to be at 4,000 metric tonnes of nuts-in-shell. After processing, these reduces to about 800 metric tonnes of marketable kernels for commercial importance; though the shells have gain their use as a bio-fuel in most homes in Kenya. Macadamia nuts production according to [2]projected to increase from the current two thousand metric ton to thirty thousand metric tons by the year 2020. With this projection, it is necessary to develop dryers that can handle the large quantity of macadamia nuts at once or that can dry same volume as present but at a shorter drying time without spoilage. This would entail increasing handling volume or reducing handling time.

[1] Directorate of Intellectual Property management and University-Industry Liaison, JKUAT University, Nairobi, Kenya
[2]Agricultural and Bio-systems Engineering Department, JKUAT University, Nairobi, Kenya
[3] Department of food science and nutrition, JKUAT University, Nairobi, Kenya

The choice of dryer should be one that produces high quality nuts for commercial viability. The quality according to [3] and [4] includes appearance of the nut (size, colour and freedom from defects), texture (firmness and crispness) and chemical value. In addition, good quality macadamia nuts should contain at least 72% oil since less than this value means that the nuts are immature and hardy. These nuts tends to over brown during roasting.

The most commonly used texture measure that is non-destructive according to [3] and [5] is the Gray-level co-occurrence matrix (GLCM). This is because GLCM features that are computed are based on the assumption that the texture information extracted from the image is found in the entire spatial relationship of pixel to its neighbors.

2. MATERIAL AND METHOD

I. Sample preparation

Two macadamia nut varieties (MRG-20 and KRG-15) Grown in Embu, Kenya were used in this study. The nuts were dehusked immediately after harvesting and transported to Juja in onion nylon bags. These were the air dried in a shade to a moisture content of 10-12.5 (d.b) for a period of two weeks.

II. Experimental set up

A Color digital camera, CDC, (model Samsung WB150F ,Samsung, South Korea), with a wide-angle lens 24 mm and high resolution of 14.2 megapixel was located vertically over the back ground at a distance of 30 cm. The camera was fixed on a static table, as shown in Fig. 1. The adjustment of the camera was standardized in manual mode with the lens aperture at f = 4.5 and speed 1/80, no zoom, no flash, intermediate resolution of the CDC (1280×720) pixels, and storage in JPEG format. The camera connected to the USB port of a PC for downloading, to analyze the digital pictures taken. The angle between the camera lens and the lighting source was 45°. Sample illuminators (Bulb light (D65000K) and the CDC was placed in a dark room to avoid the external light and reflections.

III. Data processing of imaging

This involved the capturing of images of the two varieties of macadamia nuts, KRG-15 and MRG-20.This was achieved by capturing six images from a nut and repeated three times for those particular nuts dried under the same condition for the two varieties. The nuts were subjected to seven different drying methods during this experiment. These were:-Initial condition, solar tent drying, oven drying at 50˚C, oven drying at 60˚C, oven drying at 50-60˚C, solar tent-60˚C and solar tent-MW drying. These images were fed to Imagej software, which extracted the colour features in RGB.

These were converted using the same software into gray scale imaging in order to compute the image texture features. GLCM Texture Too plugin was used in the computation of the following texture features: - Contrast, Angular Second Moment (ASM), Homogeneity and Correlation in all directions (0˚, 45˚, 90˚ and 135˚) at a certain displacement of the GLCM (d=1). This displacement [6] is the distance between two pixels whose repetition is to be examined

Figure 1: Imaging set up

3. RESULTS AND DISCUSSION

a) Statistical Texture Parameters Extraction from GLCM

I. KRG-15

Fig. 2 shows the means of the four texture parameter (contrast, angular second moment, homogeneity, and correlation) of KRG-15 dried using different drying methods in all direction (0˚, 45˚, 90˚ and 135˚). Contrast for KRG-15 nuts dried using oven dryer at 50˚C and 50-60˚C where much higher compared to the KRG-15 nuts dried using other drying method in all the directions. This meant that these drying methods triggered high local variation in KRG-15. The same conclusion observed that nuts dried using oven dryer at 50˚C and 50-60˚C were less homogeneous as compared to those dried using both solar tent dryer and solar tent-MW.

For ASM, solar tent–MW was the highest followed by Initials, solar tent dryer, solar tent-60 ˚C, oven drying at 50-60˚C and 60˚C but was least in oven drying at 50˚C for directions. This indicate that KRG-15 nuts dried using solar tent and microwave have uniform local texture distribution as compared to the other drying methods. However, the low correlation values of krg-15 nuts dried using solar tent-MW indicate that these nuts had rough texture compared with initial KRG-15 nuts before drying and those dried using solar tent dryer (highest correlation values) which are smooth in their texture. Solar tent-60˚C and oven drying at 60˚C gave nuts with good texture.

II. MRG-20

Fig. 3 shows the means of the four texture parameter (contrast, angular second moment, homogeneity, and correlation) of MRG-20 dried using different drying methods in all direction (0˚, 45˚, 90˚ and 135˚). Contrast for MRG-20 nuts dried using solar tent-MW and oven dryer at 50˚C where much higher compared to MRG-20 nuts dried using other drying method in all the directions. This means that these drying methods triggered high local variation in MRG-20 nuts. However, the nuts dried using oven dryer at solar tent-MW and 50˚C were less homogeneous as compared to those dried using solar tent dryer.

ASM value for solar tent –MW was the highest followed by Initials, solar tent dryer, oven drying at 50-60˚C, solar tent-60˚C and 60˚C but was least in oven drying at 50˚C for the direction. This indicate that MRG-20 nuts dried using solar tent and microwave have uniform local texture distribution as compared to the other drying methods. Fig 4.25, the low correlation values of MRG-20 nuts dried using solar tent-MW indicate that these nuts had rough texture compared with initial MRG-20 nuts before drying and those dried using solar tent dryer had the highest correlation values) which are smooth in their texture. Solar tent-60˚C and oven drying at 60˚C gave nuts with good texture.

Fig. 4 and 5 shows that the KRG-15and MRG-20 at initial (before drying) and those dried using solar tent dryer had the lowest contrast value at 0˚ direction. The lesser the contrast value the better the nut texture. KRG-15 dries using oven 50˚C, oven 50-60˚C and solar tent-MW had the worst texture in comparison to solar tent drying alone. However, for MRG-15, this was observed when drying using solar tent-MW.

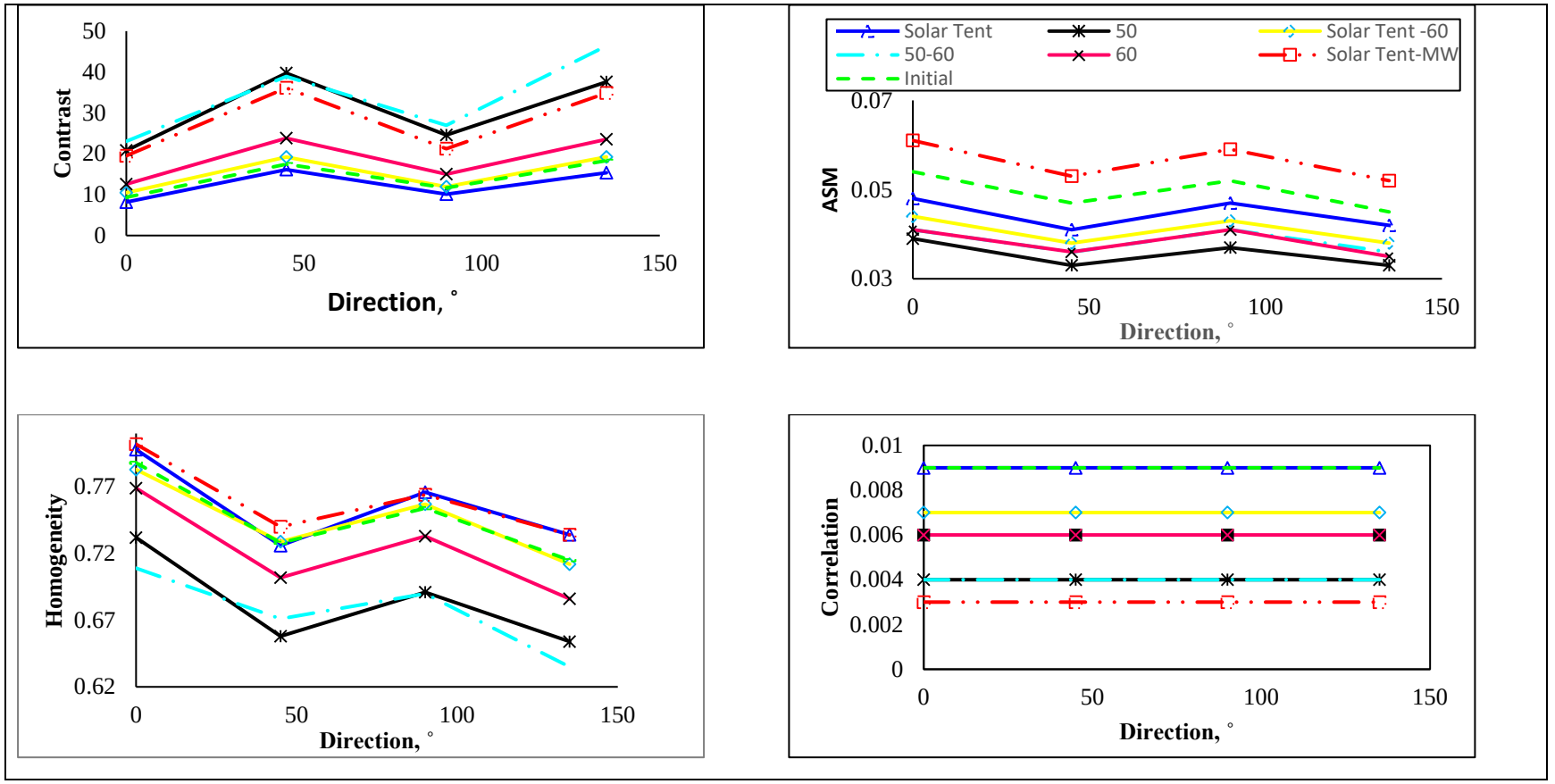


Figure 2: KRG-15 macadamia nut's Texture parameter for different drying methods using GLCM

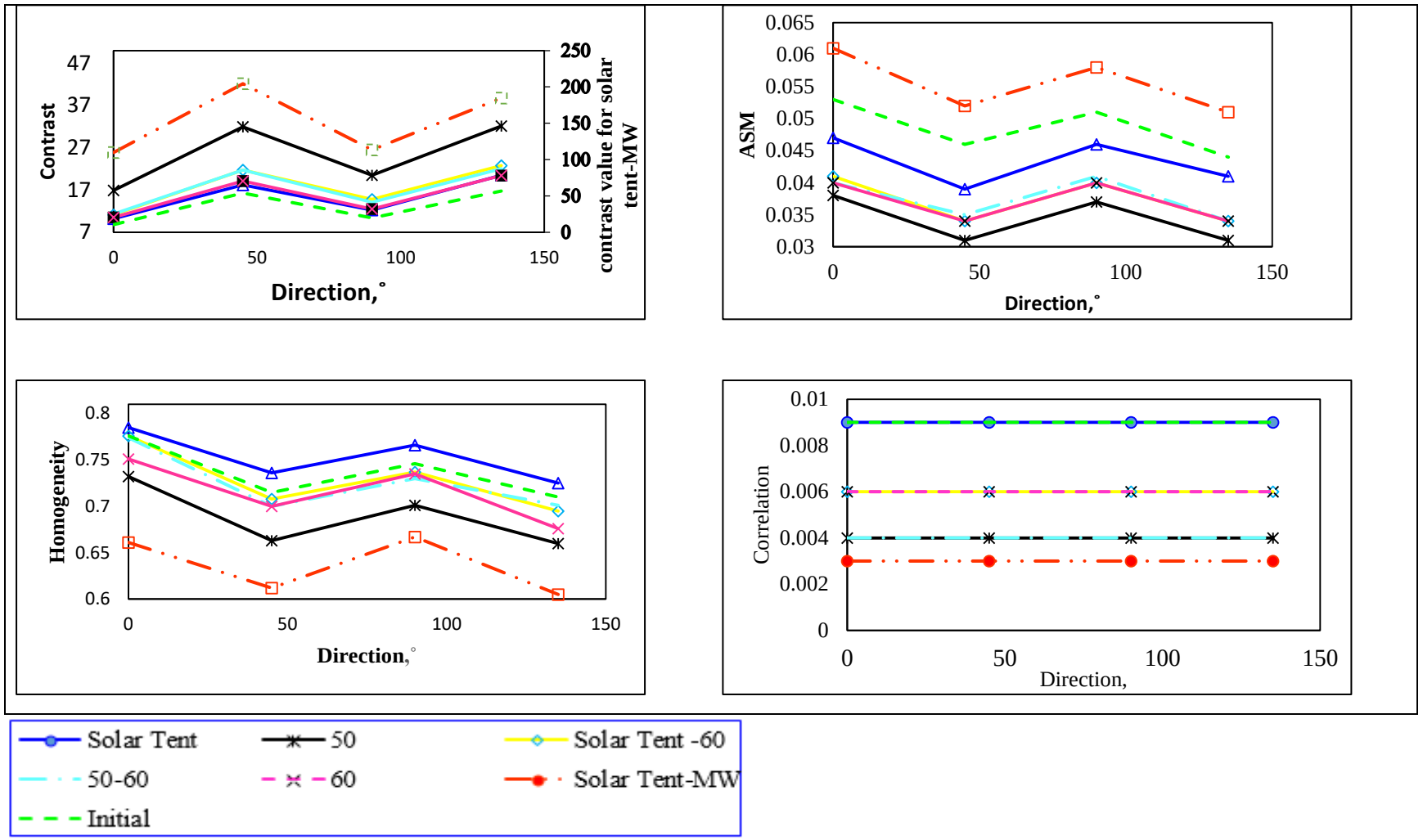


Figure 3: MRG-20 macadamia nut's Texture parameter for different drying methods using GLCM

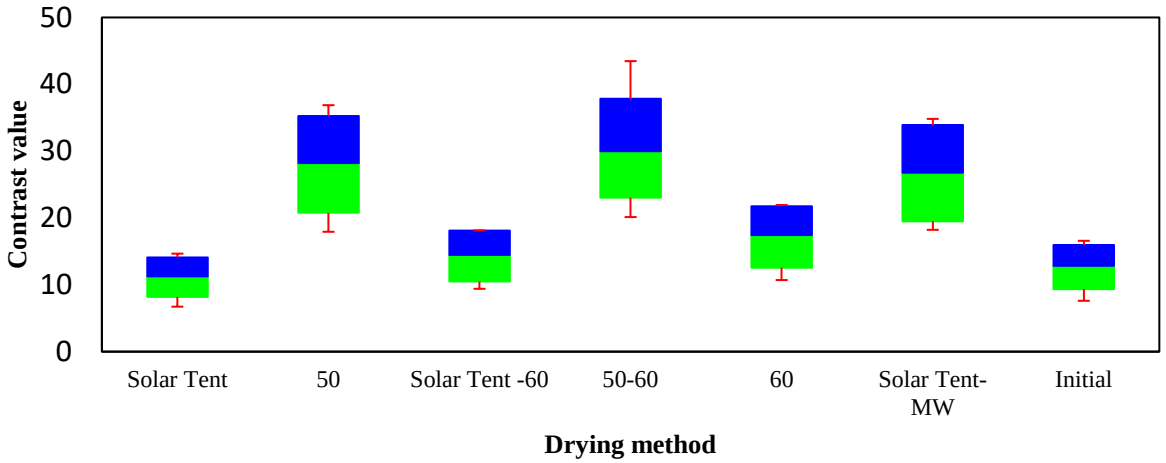


Figure 5: Contrast graph for KRG-15

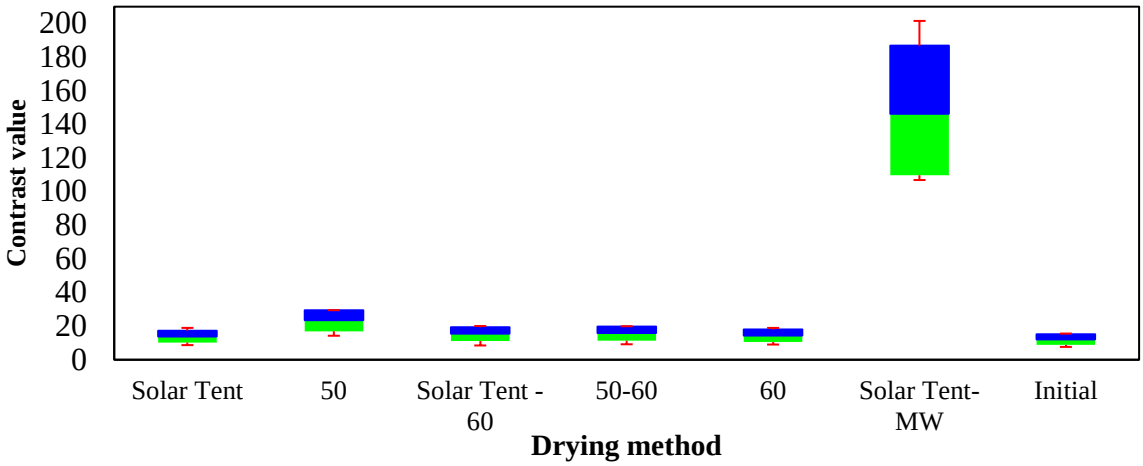


Figure 6: Contrast graph for MRG-20

b) ***Classification of drying method based on the texture parameters***

Table 1: Classification of Macadamia nuts variety at different direction

Direction	0˚	45˚	90˚	135˚	Mean
KRG-15	71%	79%	71%	71%	73%
MRG-20	82%	68%	71%	71%	73%
Mean	77%	74%	71%	71%	73%

The discrimination efficiency of the different varieties of macadamia nuts at different directions of 0˚, 45˚, 90˚ and 135˚ were found as shown in table 1. It is evident that the horizontal direction at angle 0˚ had the highest average discriminative efficiency of 77% as compared to the other directions

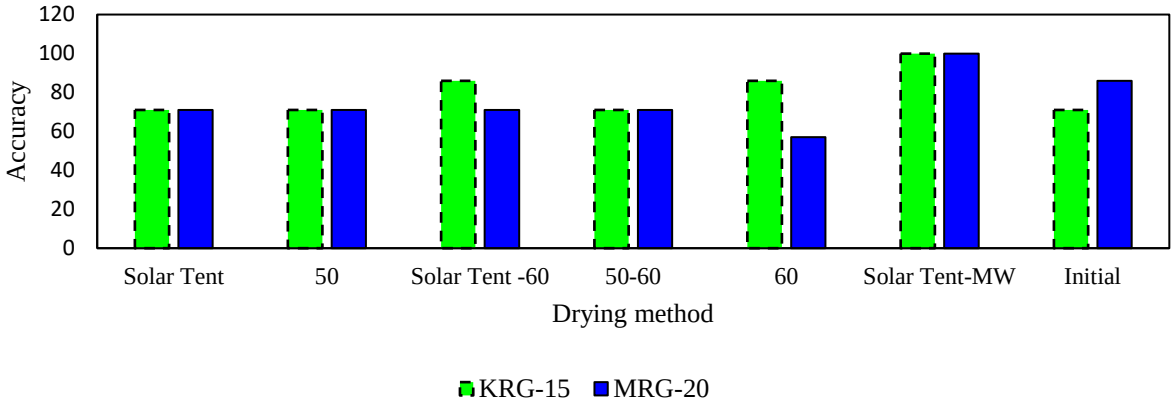


Figure 7: Classification of drying method based on the texture parameters at angle 0˚

Fig 3. Shows the classification of nuts dried using different methods at 0˚ direction. The accuracy was highest on nuts dried using solar tent-MW for both variety, KRG-15 and MRG-20.

4. CONCLUSIONS

Texture measures derived from the grey-level co-occurrence matrix (GLCM) of both KRG-15 and MRG-20 nuts images to identify the influence of drying methods on texture quality. High classification accuracy of 77% for correctly identifying influence of drying method was achieved using the GLCM features at horizontal direction at angle of 0˚, as observed by [3]. Solar tent drying method produced superior quality Macadamia nuts for the two varieties with the GLCM feature values (contrast, Homogeneity, Correlation and ASM) of 8,179, 0.048, 0.798 and 0.009 respectively for KRG-15 and 10.238, 0.789, 0.007 and 0.047 respectively for MRG-20. The dryer with the least quality in comparison to the other drying method was combination Dyer of Solar tent dryer and microwave (Solar tent -MW) for with a correlation value of 0.03 for both KRG-15 and MRG-20.

ACKNOWLEDGEMENT

The National Commission for Science, Technology and Innovation of the Republic of Kenya and Jomo Kenyatta University of Agriculture and Technology funded this study. I also want to recognize the Technical assistance from Practical training Centre (PTC) and the department of agricultural and Bio-system engineering. Finally, Mrs. Susan W. Ng'ang'a for the assistance in the collection of the nuts.

REFERENCES

[1] A. Mbora, R. Jamnadass and J.-P. Lillesø, "Growing high priority fruits and nuts in Kenya: Uses and management.," Digital Process Works ltd, Nairobi, 2008.

[2] N. M. Muthoka, P. D. N. Kiuru, J. Mbaka, A. N. Nyaga, S. J. N. Muriuki and C. N. Waturu, "Macadamia Nut Production and Research in Kenya," *The African Journal of Plant Science and Biotechnology,*, vol. 2, no. 2, pp. 46-48, 2008.

[3] G. ElMasry, N. Wang, A. ElSayed and M. Ngadi, "Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry," *Journal of Food Engineering, 81(1), 98–107,* vol. 81, no. 1, pp. 98-107, 2007.

[4] K. Chandan, C. Siddharth, A. R. Narmadha and M. g. Harika, "Classifications of Citrus Fruit Using Image Processing -GLCM Parameters," *International Conference on Communications and Signal Processing (ICCSP),* pp. 1743-1747, 2015.

[5] S. Ondimu and M. H., "Effect of probability-distance based Markovian texture extraction on discrimination in biological imaging," *Computers and Electronics in Agriculture, ,* vol. 63, no. 1, pp. 2-12, 2008.

[6] A. Gebejes and R. &Huertas, "Texture Characterization based on Grey-Level Co-occurrence Matrix," *Conference of Informatics and Management Sciences,* pp. 375-378, 2013.

A Step-By-Step Approach to Retrofit and Automation of PCB Machines Using PLC

John Nyutu Kamau, EN371-2188/2015

Msc. Student, Department of Electrical and Electronics Engineering

Jomo Kenyatta University of Agriculture and Technology, Juja, Kenya

P.O. Box 8191 – 0100 THIKA, KENYA

Email:jnyutu@gmail.com

ABSTRACT

Abstract – Retrofitting of machine has today become very popular for manufacturing and processing in small, medium and large industries as demand for high quality, greater efficiency and automated machines increase by day. It offers numerous advantages in the area of cost effectiveness and higher productivity. Technological upgradation and refurbishing of old machine restore them close to their original performance levels, contain capital costs, add value to the process and address obsolescence. Retrofitting can be achieved through a phased approach from the perspective of comprehension or through step-by-step approach to automate, understanding the needs thus employing the best control system to achieve the desired result. Simplification of engineering practices and precision control requirements of manufacturing processes by absorbing technological changes can result in significant cost savings. The most cost-effective way which can pay rich dividends in the long run is adopting flexible automation via a planned approach towards integrated control systems. It requires a conscious effort on the part of plant managers to identify areas where automation can result in better deployment/utilization of human-machine-interface and its implementation to achieve higher productivity. This paper highlights a case study of retrofitting and automation of a Del Monte's old FMC Caser Machine (Model 7 Non Shock Caser) with a Programmable Logic Controller, PLC, based control system using a step-by-step approach for a successful performance to address problem of: obsolescence, reduced productivity, increased down time, increased repairs and increased maintenance costs.

Keywords: Electro-Pneumatics, Industrial Automation, Programmable Logic Controller (PLC), Printed Circuit Board (PCB), Retrofitting of Machine

I. INTRODUCTION

Over the years the demand for high quality, greater efficiency and automated machines has increased in the industrial sector of all kinds. They require systems with high accuracy, greater flexibility, continuous monitoring and control. But now-a-day's rapid growth in technology has come-up with different solutions such as PLC which will fulfill all the requirements of the industrial processes through automation. Simplification of engineering practices and precision control requirements of manufacturing processes by absorbing technological changes can result in significant cost savings. The most cost-effective way which can pay rich dividends in the long run is adopting flexible automation via a planned approach towards integrated control systems.

Since customers today have access to world markets due to globalization and liberalization of economies, they have at liberty to trade-off with innovative, indigenous solutions or readily available expensive solutions. The retrofit is an attempt of indigenous and innovative solution to contain the capital costs and add value to the processes by mix and match of inexpensive available technological inputs to result in improved quality at a lower price. Hence market life of products can be increased by absorbing technology through retrofits to addresses obsolescence.

Upgrading the existing machines through retrofits can bring many a benefits of new equipment at a fraction of the cost. Much of the savings depends on the application. For example, if the control circuitry and selection represents 10 to 20% of the machine value, and its replacement just would give a new-machine of same performance, the machine

owner would save 80% of capital cost of new machine purchase. This justifies retrofitting the machine by changing the state-of–the-art control circuitry. Upgrading the machine for performance enhancements through retrofits is possible. These engineering solutions are ingenious and innovative which will render huge cost savings and higher productivity levels. Micro, Small and Medium Enterprises, MSME, sector will survive adapting these engineering practices wherever it necessitates to remain competitive in the era of technological revolution by remaining ingenious and innovative.

The typical examples could be reduced machine's productivity or increase in the cost of support systems would invite ingenious retrofit solutions. Another example could be increased down-time, increased repairs, increased maintenance costs and calibration costs would offer an opportunity to set right things by suitable retrofits ingeniously. Often a suitable closed loop control system can compensate for deficiency in performance [1].

Finally, a subtle factor affecting the new-versus-retrofit decision is the time needed to get a machine up and running productively and economically. It may be quicker to retrofit using off-the-shelf parts compared to typical lead times with new machine purchases. It is a good thing that innovations are happening at the start-up and MSME sector because they have to ever remain competitive.

II. STEP BY STEP APPROACH METHODOLOGY

Retrofitting refers to the addition of new technology or features to older system. Principally retrofitting describes the measures taken in the manufacturing industry to allow new or update parts to be fitted to old or outdated assemblies. It offers numerous advantages in the area of cost-saving, the optimization of existing plant components, adaptation of the plant for new or changed products, as well as in maintaining a high quality [2].

A review of the relevant literature [2][3][4][5][6], one can develop a step by step approach of typical retrofitting of machine, which consists of six steps: analyze existing design and review documentation, create a new design or re-engineering design, develop Automated control program and simulate to verify functionality, build the hardware system as per the design by proper selection of equipment, validate the system through testing and demonstration for the desired automated control operation of the machine, and generate complete technical data package and manuals to support operation of the machine. All of which can be represented by the block diagram shown in Fig. 1.

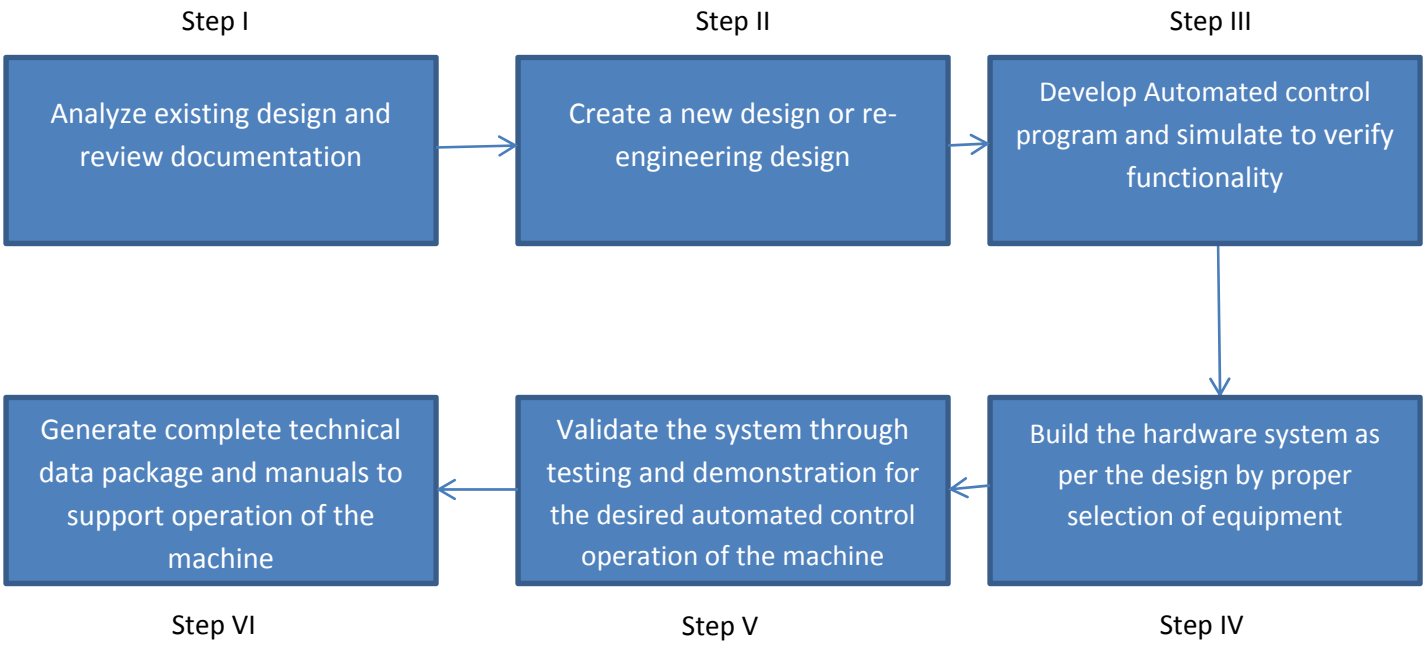


Fig.1: Step by Step Approach for retrofitting

III. AUTOMATION AND MODERN TECHNOLOGY ADVANTAGES

Industrial automation uses computers and controllers to control industrial machinery and processes to optimize productivity and delivery of services. Automation greatly decreases the need for human sensory and mental requirements [7]. The impact of automation in industries is as follows:

- Increase productivity and reduce cost.
- Emphasis on flexibility and convertibility of manufacturing process. Automation is now often applied primarily to increase quality in the manufacturing process, where automation can increase quality substantially.
- Increased consistency of output.
- Replacing humans in tasks carried out in hazardous environment.

In the mechanization and automation of industrial processes, pneumatics has gained great significance because of the easy implementation of often-needed straight-line, back and forward motion using pneumatic cylinders; swivel motion and rotation using rotary drives; revolving cylinder engines and stepping drives [8]. Until a few years ago the compressibility of the energy carrier, i.e. compressed air, basically limited the automation of fast motion sequences to simple adjusting movements where stop points were implemented by mechanical limit stops. Electro pneumatic controls combine the best features of electronic and pneumatic controls. Such system consists of pneumatically actuated valves, electrical/electronic controllers, sensors and control systems Electro-pneumatics is used in most areas of industrial automation. Production, assembly and packaging systems worldwide are driven by electro-pneumatic controls. Electro pneumatic controllers have the following advantages over pneumatic control systems:

- Higher reliability.
- Lower planning and commissioning effort for complex controls.
- Lower installation effort.
- Simple exchange of information between several controllers

Considering the varied demand and increasing competition, one has to provide for flexible manufacturing process. One of the latest techniques in solid state controls that offers flexible and efficient operation to the user is “Programmable Logic Controller”. The basic idea behind these programmable controllers was to provide means to eliminate high cost associated with inflexible, conventional relay controlled systems. Programmable controllers offer a system with computer flexibility. Programmable Logic Control or PLC as it is universally called is the ‘work horse’ of industrial automation. It is important because all production processes go through a fixed repetitive sequence of operations that involve logical steps and decisions. A PLC is used to control, time and regulate the sequence [9].

IV. CASE STUDY

4.1 Overview of Project

The aim of the project was to retrofit a Del Monte’s old FMC Caser Machine (Model 7 Non Shock Caser) using PLC. During the preliminary phase of project, the machine controller was found to be based on an old and inflexible Printed Circuit Board, PCB, board. The board was marred with frequent breakdown and its repair was very expensive. The manufacturer, FMC, had termed the board obsolete and with no replacement. The manufacturer recommended replacement of the entire machine with a new one. Most of the pneumatic cylinders and electro-pneumatic (solenoid) valves in the machine were in good condition and required no replacement. The machine controller was to be replaced by the PLC. Also the whole automatic operation of the machine was to be maintained. With these considerations, the main objective set for the project was to design, develop and implement automated controller for the machine in order to upgrade the technology. A step by step approach was to be used to achieve the objective.

4.2 Method and Implementation

Applying the proposed approach is to take each step of retrofitting in questioning and assessment through the systematic process.

4.2.1 Step I: analyze existing design and review document

The objective of this step is to analyze existing design and reviews documentation. The machine electrical components are reviewed and their location identified and their electrical specification analyzed. Electrical and logical schematic is also reviewed and number/nature of inputs and outputs explored and investigated. The fig.2 below shows the caser machine electrical components that was reviewed in details and targeted in the retrofitting.

The electrical schematic shown in fig.3 shows the PCB inputs and outputs that were used to design the new PLC controller and determine electrical sequence of operation.

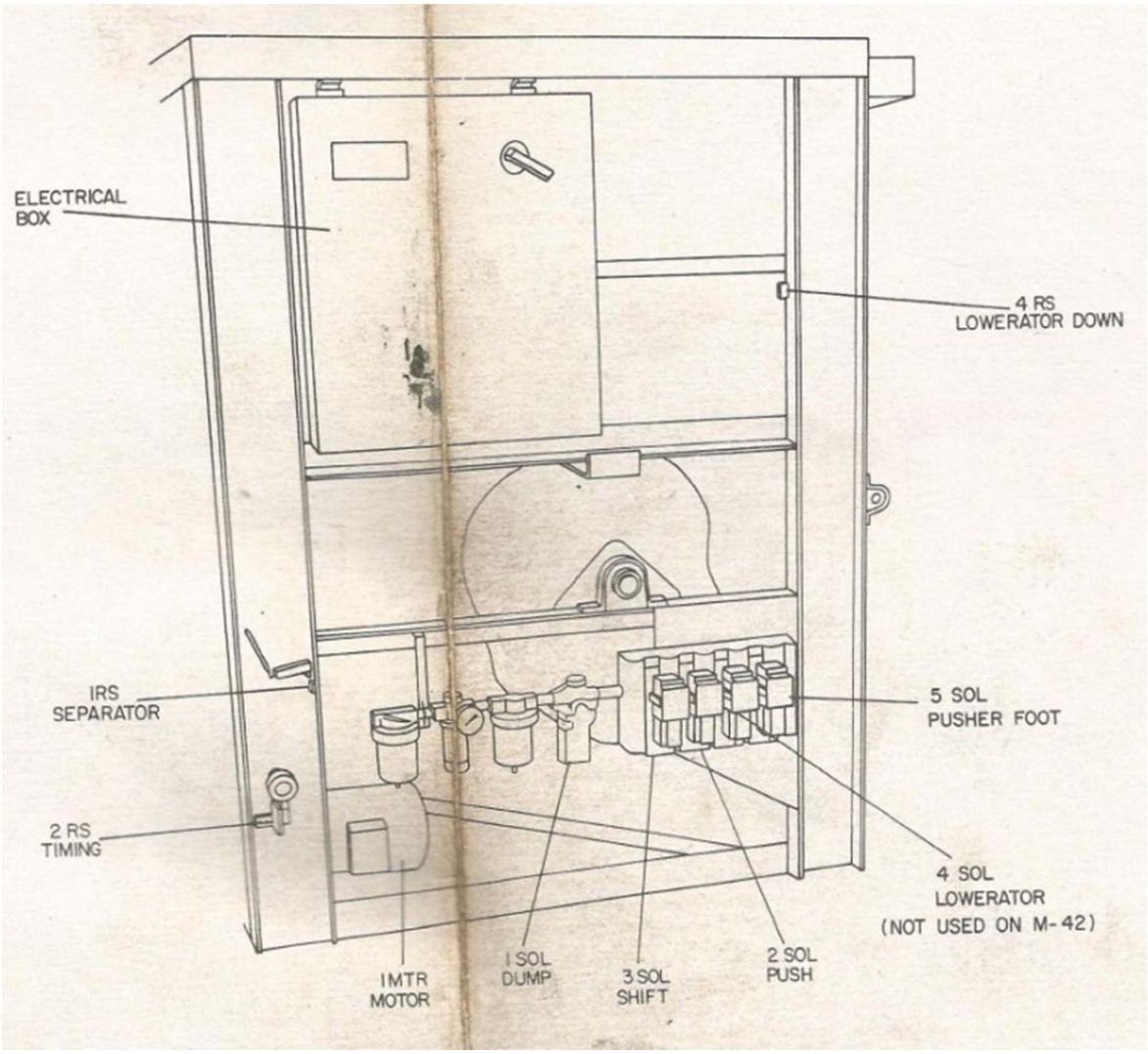


Fig.2: **Electrical component locations**

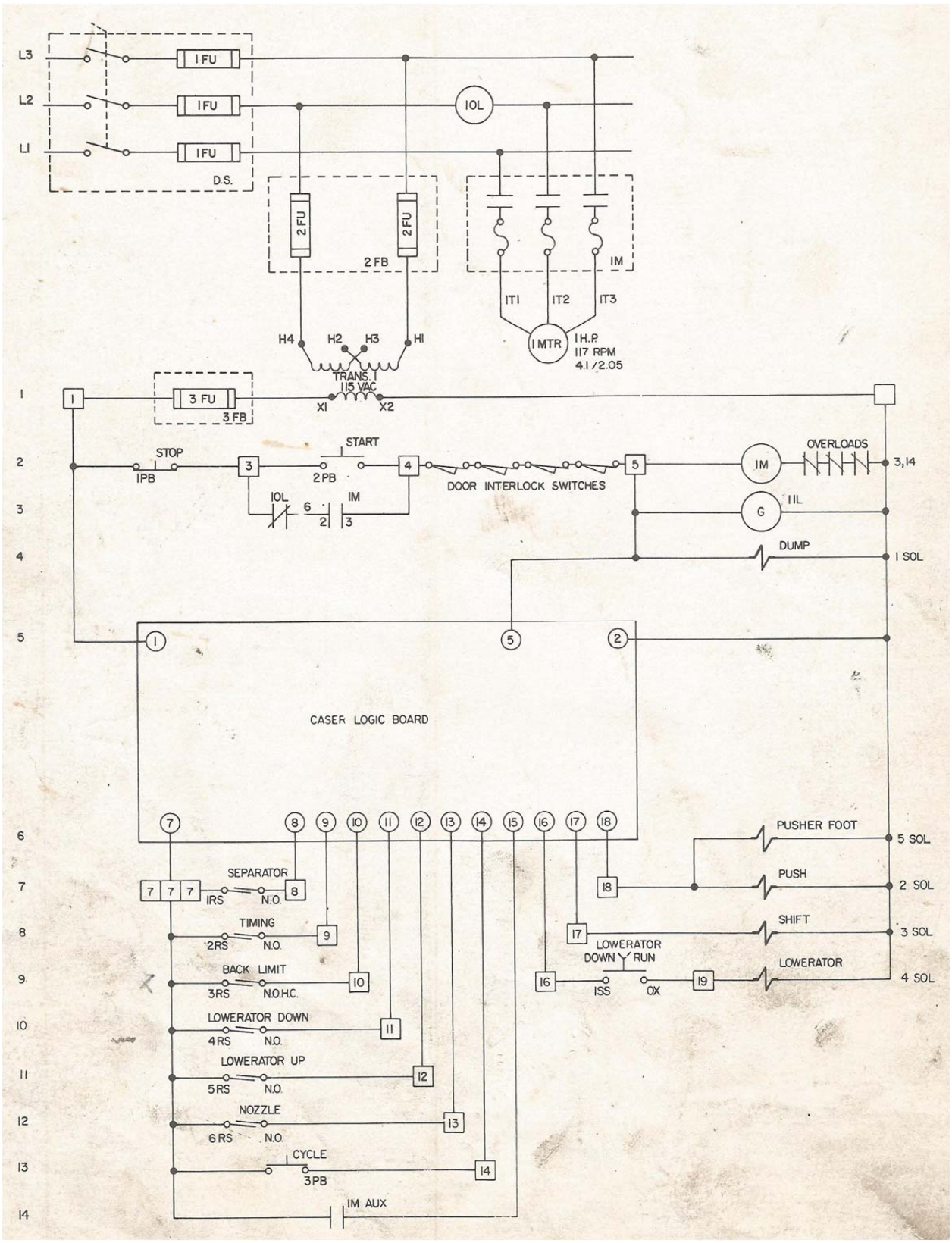


Fig.3: Electrical schematic

4.2.2 Step II: create new design or re-engineering design

Problem presented in this step is how to replace the PCB logic board with a PLC to control the machine. The purpose of this step is creates new design or re-engineering design. The existing machine is driven by a number of solenoid valves, limit switches, push button switches, and accessories. This is used in determining the number of input and output for the design of control by PLC to replace the old system. After a detailed exploration and investigation, it is found that 8 inputs and 3 outputs are required. In this stage, a PLC control is designed. The fig. 4 below show the new PLC design showing input/output relays, protection MCB and siemens Logo PLC.

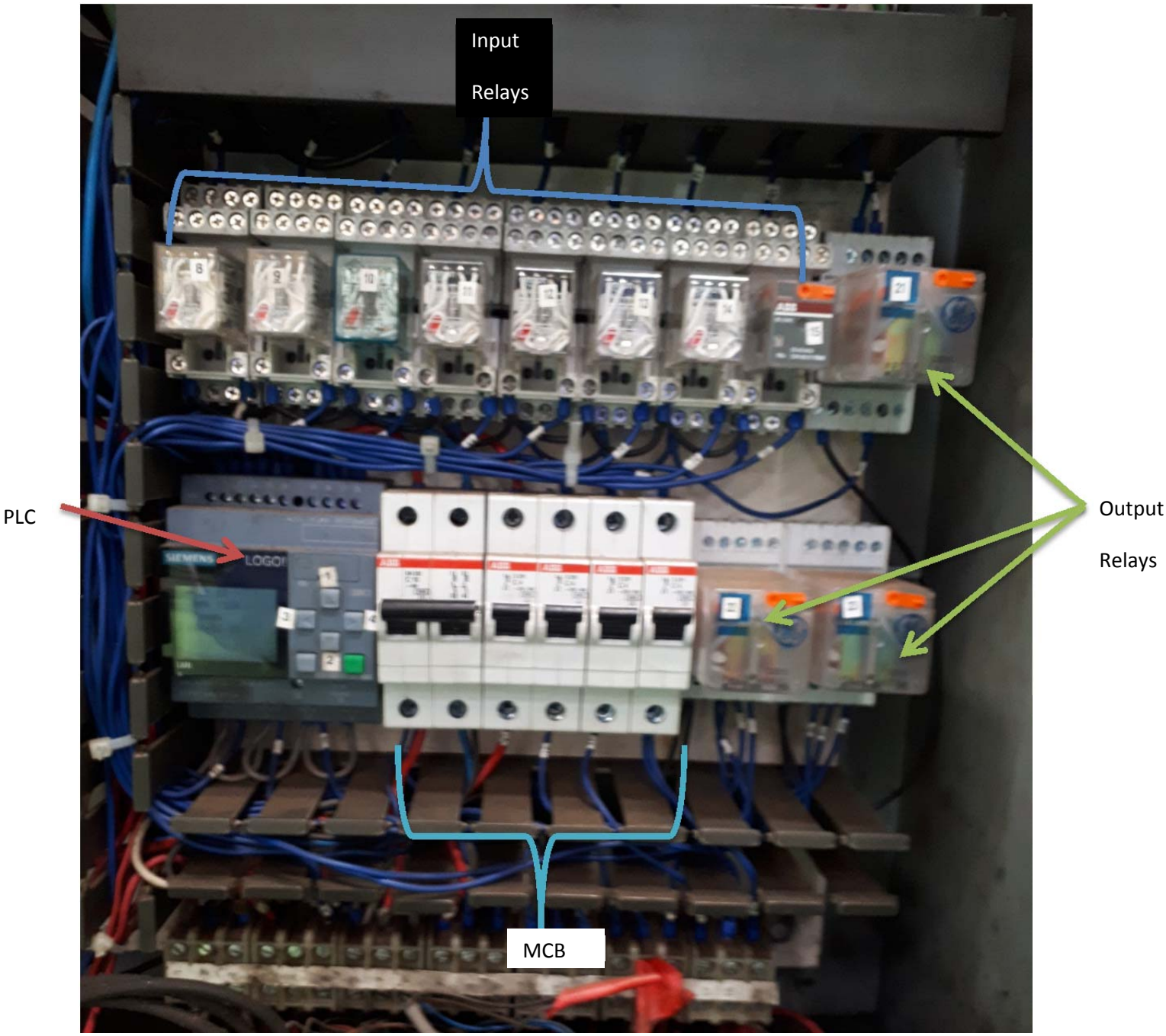


Fig.4: new design

4.2.3 Step III: Develop automated control program and simulate to verify functionality

Problem presented in this step is how one would know the machine functions as it should. The objective of this step is to develop a PLC program and simulate it to verify functionality. The Siemens logo software comfort version 8.1 was used to simulate the program. By simulating manual cycling as given by the manufacturer, one would ascertain a perfect electrical sequence of operation and functionality. The manual cycling is performed as follows: pushing cycle push button (3PB) with the nozzle reed switch (6RS) actuated, will cycle machine at any time. Without 6RS actuated, the cycle button will only cycle machine until push sequence is scheduled to occur. At this point 6RS must be actuated to make the cycle button operative.

4.2.4 Step IV: Build the hardware system as per the design by proper selection of equipment

Problem presented in this step is how would you know that system that you designed is workable reality. The objective of this step is building the control panel to verify design. In the hardware part, overall component such as PLC, relays, transformer and switchgear circuit will be integrated to form the complete control circuit. The detailed list of all the hardware used in the project is listed in Table 1 below:

Table 1. List of hardware used in the project

NO	**Hardware Description**	**Make**	**Rating**	**Qty**
1	LOGO! 8 Logic Module	Siemens	230 Vac, 8 x Input, 4 x Output With Display	1
2	11 PIN RELAY	ABB	120VAC, 3PCO, 10A	3
3	14 PIN RELAY+LED	ABB	24VAC, 10A 4PCO	8
4	Miniature Circuit Breaker S201-K1	ABB	1P K 1A 480Y/277 SUPP	3
5	Miniature Circuit Breaker S201-K4	ABB	1P K 4A 480Y/277 SUPP	1
6	Miniature Circuit Breaker M202-10A	ABB	M202- 10A MCB OM 10A,2P	1
7	Control transformer	RTR energia	200VA, 50Hz, 0-240V, 0-24v	1
8	Control transformer	RTR energia	100VA, 50Hz, 0-240V, 0-110v	1

4.2.5 Step V: validate the system through testing and demonstration for the desired automated control operation of the machine

Problem presented in this step is how would you know that the machine is running according to function correctly. The purpose of this step is to validate through testing and demonstration of machine. The machine was tested through the entire electrical sequence of operation.

Electrical sequence of operation (as shown in fig.3):

- Depressing start button (2PB), energizes 1M motor starter through 1PB, 2PB and the four door interlock switches. If any of the door interlock switches are not closed, 1M cannot energize and machine will not start. When 1M starter energizes, an auxiliary contact of 1M will close energizing the 1M input of the caser logic board. When the 1M input of the logic board is energized, the remaining inputs and the three outputs will become operative.
- The 1M input must be energized for the logic board to function.
- Placing a case on the nozzle actuates the lowerator up reed switch (5RS). When 5RS is actuated, the caser logic board energizes the lowerator air solenoid (4SOL) causing the lowerator chair to rise. As lowerator rises, the nozzle reed switch (6RS) is actuated.
- As incoming cans fill the individual lanes of the caser, the separator reed switch (1RS) actuates. The caser logic board will energize the shifter air solenoid (3SOL), when both 1RS and 6RS are actuated. When 3SOL is energized, the shifter assembly blocks cans entering the separator and actuates an air limit switch mounted above the shifter.
- Actuating this air limit switch causes the main drive shaft to rotate by triggering a single revolution wrap spring clutch brake. As main drive shaft rotates, the timing reed switch (2RS) is actuated momentarily.
- When 2RS actuates, the logic board will de-energize the shift air solenoid (3SOL) after an adjustable time delay. A logic board mounted potentiometer marked “SHIFT RETURN” controls the amount of time the shifter assembly remains in the shifted position after 2RS is actuated. This time is adjustable from 0 to 0.4 seconds.
- The timing reed switch (2RS) also registers one count in the internal counter of the caser logic board. The logic board will energize the push air solenoids (2SOL, 5SOL), when 2RS actuates.
- Energizing 2SOL and 5SOL causes the pusher feet to drop and the pusher cylinder to extend. As the pusher cylinder extends, the lowerator down reed switch (4RS) is actuated. When 4RS actuates, the logic board will immediately de-energize 4SOL and after a time delay, will de-energize 2SOL and 5SOL. This time delay is adjustable by means of a potentiometer marked “FORWARD LIMIT”.
- This potentiometer is mounted on the caser logic board and is adjustable from 0 to 0.2 seconds. When lowerator air solenoid (4SOL) de-energizes, the filled case is lowered from nozzle area. When push air solenoids (2SOL,5SOL) de-energizes, pusher feet raise and the pusher cylinder retracts. When the pusher cylinder is fully retracted, the back limit reed switch (3RS) actuates. The back limit reed switch (3RS) must be actuated before the next cycle may begin.
- In 2,3,4 and 5 tier operation, the sequence is identical to one tier with the exception of the push cycle. The push cycle will not occur until a sufficient number of tiers have been lifted to the tiering chamber. This is accomplished by the internal counter of the caser logic board.

The machine was run in production process for three months. During this time the PLC program was modified many times in response to the user feedback.

4.6 Step VI: generate complete technical data package and manuals to support operation of the machine

The problem presented in this final step is how the operator and maintenance team would use/maintain this machine. The purpose of this step is to generate complete technical data package and manuals to support successful operation and maintenance of the machine.

V. RESULTS AND DISCUSSION

5.1 Retrofitting of machine

The results of retrofit show that engineers in the industry can successfully replace PCB controller by a PLC and upgrade an obsolete machine to give it a new-machine performance at fraction of its cost.

5.2 Production achievement

Once the machine was retrofitted and objective achieved, the company addressed the problem of obsolescence, reduced productivity, increased down time, increased repairs and increased maintenance costs. It was costing the company kshs.200,000 per month on outsourced repair of control PCB board and

kshs.500,000 on loss of production. The retrofit only cost kshs.100,000 and has been working well for the last six months. Thus the retrofit and automation of the machine paved the way for increasing the productivity and reliability, time saving and PLC display enables the operation of machine easier for non-skilled labors too. Also the machine operation had good quality of product. As the project focuses on the closed loop control of the process, this will fulfill all the requirements of the industry with good levels of accuracy and repeatability thereby yielding a more robust system.

VI. FUTURE WORK

Since the benefit acrued with the retrofit is immese, it is therefore recommended that the retrofit be applied to the other two caser machines, which are still running on the old PCB board.

REFERENCES

[1] J.Hyung, Kim Harms, R. Seliger and G. Dept. of Assembly Technol. & Factory Manage., Tech. Univ. Berlin, “Automatic Control Sequence Generation for a Hybrid Disassembly System”, IEEE Transaction on Automatic Science and Engineering, Volume: 4, Issue: 2 on page(s): 194-205**.**

[2] Possibilities and Advantages of Retrofit, Sitola GmbH & Co. KG, (2008).

[3] The Retrofit Advantage, FANUC FA America, (2011).

[4] Custom Retrofitted Machinery to Full Automation, http://www.rebuildersunlimited.com

[5] K. Ivan, K. Marian, and J. Frantisek, “Reverse engineering as an education tool in computer science,” in Proc. the 9thIEEE Conf. on Emerging eLearning Technologies and Applications, pp. 123-126 (2011).

[6] S. Frederic, M. A. Nicolas, and V. Philippe, “Collaborative reverse engineering design experiment using PLM solution,” Int. J. of Engineering Education, vol. 27, no. 5, pp. 1037-1045 (2011).

[7] O.Masao and H.Yutaka, “Recent Developments on PC+PLC based Control Systems for Beer Brewery Process Automation Applications”, SICE-ICASE International Joint Conference 2006 Oct. 18-21, 2006 in Bexco, Busan,.

[8] A.Gabor, Biacs, S. Milan and Adzi “PLC Control for a Rotating Ironing Press**”,** Intelligent Systems and Informatics, 2008. SISY 2008. 6th International Symposium on Publication Date: 26-27 Sept. 2008.

[9] J. J. Harris, J. D. Broesch, and R. M. Coon, “A combined PLC and CPU approach to multiprocessor control,” *in Proc. 16th IEEE/NPSS Symp.Fusion Engineering*, vol. 2, 1995, pp. 874–877.

A Review on Optimal Network Reconfiguration in the Radial Distribution System using Optimization Techniques

Juma S A, Ngoo L M and Muriithi C M

***Abstract*—**Distribution network systems face many problems, mostly of enormous power losses compared to the rest of the network, which lead to poor performance and degradation of the system and components. For maximum benefit of distributed generation (DG), optimal placement and sizing integrated in the radial distribution system has to be performed to obtain minimum power losses and improve the voltage profile of the system. The contribution of optimum network reconfiguration has greatly benefited in effective power loss reduction. The changes in network topology alters the current flowing through the line and the resistive losses. To solve network reconfiguration problems in radial distribution systems, strategic optimal network reconfiguration which can minimize power losses while maintaining the operating constraints is used to review various optimization techniques.

***Keywords*—**Network reconfiguration, Power loss reduction, Radial distribution system, Optimization techniques.

I. INTRODUCTION

The electric utility restructuring has become a major concern on the overall system reliability as the industry expands in size, becoming increasingly complex. The complexity of the electric power systems has contributed to some of the technical and economic problems, which power system engineers solve by engaging in effective planning activities. The distribution lines are normally forced to operate at low voltages to enable smooth transition of power to the consumer loads. Due to the low voltages and high currents in the distribution system, high power losses in the form of heat (I^2R) are experienced, making the distribution system less efficient. The distribution networks normally have radial feeders, which are rigid and vulnerable to outages with over-loading experienced in many parts of the system. It is essential to look at the methods of reduction in distribution network losses when planning to lower energy consumption, cost and balancing load-generation power. Reduced power losses will increase the life span of equipment due to enhanced power quality and reliability of the network. Feeder reconfiguration is not only limited to reducing power losses, but also benefit in system security, improving the voltage profile, load balancing of the system and improve in efficient use of DG systems.

S. A. Juma, Department of Electrical Engineering, PAUISTI (JKUAT) (phone: +2540797034602; e-mail: shaibu900@gmail.com).

The topological arrangement of the distribution feeders is manipulated by varying the tie and sectionalizing switches while maintaining the constraints levels known as Distribution Network Reconfiguration (DNR) [1][2][3][4]. The reliability of a distribution system is of great importance as electric power has to be supplied to the consumer in an elastic manner in case of any contingencies while maintaining protection of equipment and feeders. The reconfiguration of the network system ensures that all network operations are carried out in lucid and most favorable conditions while maintaining adequate levels that are reliable and secure for quality power supply [5][6].

From the recent studies on network reconfiguration, numerous proposed techniques were applied to find an optimum solution to power loss reduction. This paper will review some of the strengths and limitations of the recent optimization techniques used in solving network reconfiguration problems in radial distribution systems.

II. DISTRIBUTION SYSTEMS

The electrical power distribution system provides power at very low voltages to the consumer loads. The two common types of distribution networks are radial distribution system, which has separate feeders that radiate from a single substation and ring main distribution system, which has feeders in a ring fashion that terminates back to the substation.

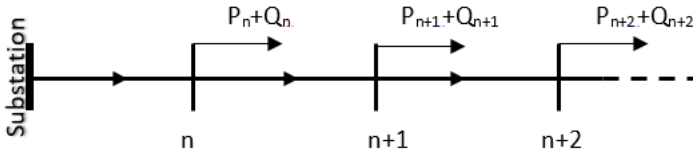


Fig. 1: A single line diagram of distribution system

The ring main system is generally expensive compared to the radial system as it has more switches and conductors required in the construction and for the power generation at such low voltages, thus a radial system is a more preferred choice [7][8].

L. M. Ngoo, Department of Electrical Engineering, MMU (e-mail: livingngoo@gmail.com).
C. M. Muriithi, School of Engineering and Technology, MUT (e-mail: cmainamuriithi@gmail.com).

However, the radial distribution system is less reliable due to its structure that has each bus connected to the source via a single path.

III. PROBLEM FORMULATION AND CONSTRAINTS

Switch state changes will manipulate the distribution network's topography and allow for the distribution of loads to be balanced accordingly and avert the system from overloading. Power loss of any line between buses in a distribution system can be evaluated by [9][10][11]:

$$P_{loss} = \sum_{n=1}^{NL} r_n \left| I_n \right|^2 \quad (1)$$

P_{loss}: the total power loss in the network distribution.
NL: Set of branches.
r_n: the resistance in the branch n.
I_n: the current in the branch n.

The most common constraints taken into account during optimization of network reconfiguration are [12]:

$$V_{n,\min} \le V_n \le V_{n,\max} \quad (2)$$

The voltage limits must be retained within the range at the buses and V_n is the voltage magnitude at *nth* bus.

Feeder's capability must have power limits in n branch:

$$k_n / P_n / P_{n\max} \quad n \in NL \quad (3)$$

$$k_n / \mathrm{Q}_n / \mathrm{Q}_{n\max} \quad n \in NL \quad (4)$$

$$\varphi(k) = 0 \quad (5)$$

Q_n ,P_n : reactive power and real power in the *nth* branch.
K_n : status topology of the branches (if branch n is closed k_n =1 and if it is open it is 0).

The topological radial structure constraints for each candidate is represented in (5). There will be no isolated nodes and the final configuration must be radial with all loads connected.

IV. METHODS OF NETWORK RECONFIGURATION WITH OPTIMIZATION TECHNIQUES

Different methods have been developed for the reconfiguration of the distribution network with some of the works including a multi-objective function for capacitor placement and DG (sizing and placement) and FACTS devices playing a role in solving network problems. This paper will summarize some of the optimization methods used in network reconfiguration of the distribution system.

Computer Intelligence (CI) is a branch of Artificial Intelligence (AI) that came up in the early 90s based on nature inspired computational paradigms and collection of numerical information [13]. Fuzzy logic (FL), Artificial Neural Network (ANN) and bio-inspired algorithms such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Tabu Search (TS), Ant Lion Optimization (ALO) and Ant Colony Optimization (ACO) are some of the AIs primarily focused on computer-based intelligence. According to Abedinia *et al.*, [14] metaheuristic optimization techniques solve real-world problems which are too complex to solve using numerical optimization methods. Some of the advantages of AI are high flexibility, simple and efficient to model. Heuristic algorithms are inspired by nature and observes the special ability of animals' superiority over other animals. They have a greedy approach that makes them easy to implement for high searching efficiency but have a drawback on convergence to a global solution when applied in a large-scale distribution system. A correct problem function must be formulated for the technique to be very effective in finding the global optimal solution. A direct search to global optimal can easily be processed in AI algorithms [1][15]. When an optimization procedure is constrained, it will mathematically determine optimal allocation of scarce resources subject to a set of constraints. Fig. 2 shows the flow chart for reconfiguration process. The optimization techniques are used to identify the best possible switching sequence of the distributed system.

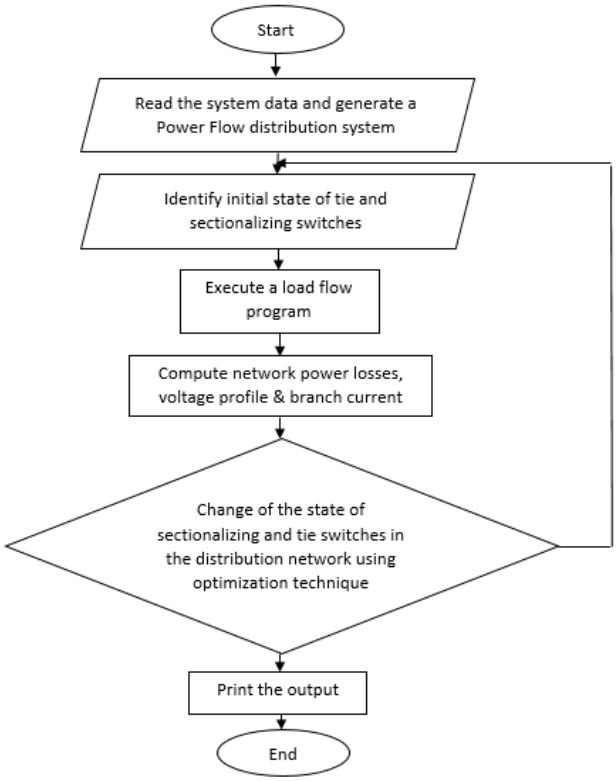


Fig. 2: Flow chart for reconfiguration process

In a Cuckoo Search Algorithm (CSA) the cuckoo birds produce and deposits their eggs in the nests of other host birds of different species and benefit by tricking the birds into brooding their eggs. CSA is an efficient population-based heuristic Evolutionary Algorithm (EA) for solving optimization problems with the advantages of simple implementation procedure and a few control parameters [1]. It is proposed to optimize the Radial Distribution Network (RDN) problem concerning power transmission losses and improving voltage profiles. The net power loss is calculated by equation (6).

$$\Delta P_{loss}^{R} = \frac{P_{loss}^{rec}}{P_{loss}^{0}} \quad (6)$$

where $P_{\text{loss}}^{\text{R}}$ is the net power loss; $P_{\text{loss}}^{\text{rec}}$ and P_{loss}^{0} is system's total active power loss after and before reconfiguration, respectively.

The initialized population of the nests will determine the nests to be checked by the algorithm in the radial topology and the load flow to be run. The fitness of each nest is calculated by the objective function. The quality of the every new cuckoo eggs produced by Levy flights will replace the bad nests from their position. The CSA was implemented for RDN by considering process of each radial structure of the network as a host nest and each egg was a solution randomly generated during initialization. The algorithm was performed on three different IEEE test systems (33-node, 69-node and 119-node) and it can be noticed how CSA is more applicable on large scale RDN due to its more noticeable effect on the 119 node test systems. CSA obtained the global optimum faster than Continuous Genetic Algorithm (CGA) and PSO.

Bradan *et al*., [16] proposed a technique which determines the optimal sequence path of switching operations given an optimal RDN and DG output with variable load using the Firefly Algorithm (FA) for daily basis power loss reduction while improving the voltage profile. FA is a metaheuristic optimization method based on the behavior of fireflies characterized by each individual in a colony having its own activity plan though the group as a whole appears to be highly organized. The optimal switching sequence path will obtain a large number of possible sequences. The first step was to determine the DGs real power output and network reconfiguration with variable load simultaneously while second step looked at determining the optimal switching sequence path to change the network configuration from the original form to the optimal form at any hour, from the obtained results in step 1. FA was applied on both the steps with the objective function of total fitness described below:

$$Min \rightarrow F = \sum_{h}^{T} \left(w_1 \times P_{loss}^{R} + w_2 \times si \right) \quad (7)$$

where h is the taken current time taken; T is the total hours (in the time frame); w_1 and w_2 are the weighting factors (w_1=w_2=0.5) and si is the voltage stability index.

Application of the firefly algorithm to achieve the distribution minimum main fitness gave an approach that had high quality and capable of realizing the optimal switching sequence path, optimal network configuration, and DG output. The proposed method indicated a possibility of being implemented in real systems with DGs.

In [17] a Hybrid Genetic Algorithm Particle Swarm Optimization (HGAPSO) was proposed, a combination of two algorithms to effectively minimize power losses and improve the voltage profile. The combined algorithm had an added advantage and combined merits for both algorithms breaking through the limitations of PSO and having an ability to control convergence and algorithmic simplicity. The GA updated positioning for the global best updates in order to avoid premature convergence experienced in PSO. Local maximum is avoided by introducing mutation by GA to PSO. The number of iterations were equal for both the algorithms with first half run by GA. The obtained solution became the initial population of the PSO and ran the remaining iterations in PSO to find the appropriate topology of the network. This paper vehemently looked at the gain of network configuration with and without DG units presence further assisting to obtain further optimum power loss reduction of the network. Despite the considerable decrease in power losses when DG units are placed, it was noted that some nodes violated voltage constraints, thus network reconfiguration also amended voltage violations to remain within acceptable limits.

Rao and Reddy [18], proposed a method which used a GA optimization approach to get optimal switching scheme for the reconfiguration of the feeder distribution network. The GA involves an initial population, selection, from the population, crossover and then mutation in the form of binary strings which represent the switch sequences in a given configuration topology. The feasible solutions should ensure that they are within the set constraints limits known as legal points. The fitness assignment to a chromosome should be chosen appropriately so that incase of any infeasible strings in the obtained solutions such a fitness function will be declined by using equation (8) if a low fitness value is experienced.

$$Fitness = 1/(1+P_c) \quad (8)$$

P_c is a high value penalty term

A roulette wheel method is used when selecting a mating population. A solution point with a bigger fitness value has a higher chance of being selected to be in the mating population. Mutation will keep the local optimum value from premature convergence. The chromosome with the highest fitness in the new generation is then decoded to give the required solution in network reconfiguration. Since the lines connected between the substations to the loads experience major losses in the copper losses, GA solved the configuration that gave minimum losses and a best voltage profile in the solution. The proposed method was tested on a 3-feeder and an IEEE 33 bus distribution system. From the results obtained it showed that the efficiency levels and voltage regulations operated better above 0.9 p.u and had reduced power line losses in the distribution network. The GA method was able to yield a global optimum solution. Proper selection of the parameters and fitness assignment must be observed carefully since they play a major role in ensuring that an optimal solution is obtained when using GA. However, an improvement can be made when performing the fitness function process during the load flow analysis. An ANN should be trained to have data for all feasible network configurations to avoid running the load flow for each configuration obtained before achieving the global optimum. GA takes quite a number of iterations to achieve an optimum solution and a high computational time which can be improved by using PSO.

In [19], the researchers proposed a modified PSO (MPSO) which accelerates the algorithm to reach optimum solution faster in network reconfiguration problem. It has a filtered random selective search space for initial positions. This

improves the search capability of the particles since unfeasible particle will be neglected, reducing on computational time and increasing on the convergence rate. PSO techniques are based on the animal behavior of swarming into a group to achieve a certain task such as movement of fish, birds flying in a flock and ants in a colony [20]. The behavior of such a biological system can be interpreted through mathematical equations to show the particle movement as it adjusts its position and velocity based on its experience and neighboring particles through a search space in order to obtain an optimal solution. IEEE 33-bus system was used to test the proposed algorithm in Interactive Power System Analysis (IPSA) software tool and the algorithm was created in python programming language. The results were compared to other versions of PSO algorithms such as Binary PSO (BPSO) and Selective PSO (SPSO).

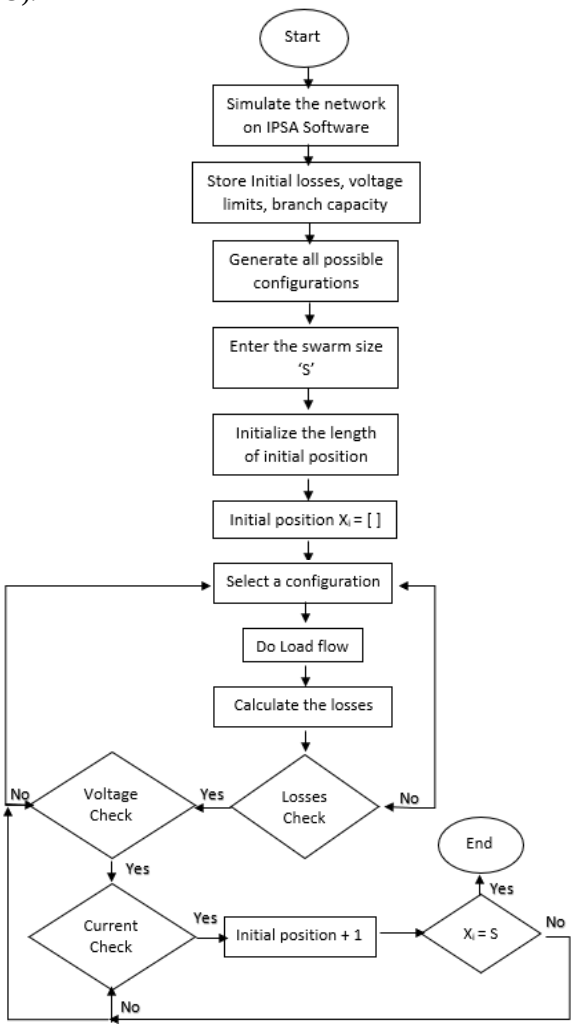


Fig. 3: MPSO Procedure in network reconfiguration [19]

Fig. 3 shows the filtered random selection adopted in controlling the initial positions of the particles. A position control algorithm will ensure that a feasible search space is maintained. For accuracy in the algorithm, the software has to be initialized after each and every loss calculation for each position. The proposed MPSO performed better in both power losses and computational time thus the modifications made achieved some improvements. The challenge faced in the algorithm was in duplication of some switches in the same position due to retaining of particles in the search space, which would violate the condition of the tie switch numbers. This might lead to the network topology to not remain radial. Lengthy computational time due to inaccuracy in results was noted when the algorithm changes the position without a reference configuration. Thus, a known configuration has to be chosen in the first attempt before each trial.

In [21], a hybridized Evolutionary PSO (EPSO) was introduced into solving the network reconfiguration problem. It is a combination of Evolutionary Programming (EP) and PSO, which will enable the algorithm to find the optimal solution with less computational time. PSO alone had less optimal solutions for the computational time in obtaining power loss reduction thus EP assisted with combination and selection method that made it faster to reach the global optimal solution. The proposed algorithm was evaluated on the IEEE 33-bus distribution system and compared to EP and PSO. The computational time was 12.2s faster than the other two methods and obtained minimum power losses achieved by final switch sequence (11, 28, 32, 33, and 34) for the open switches. The total losses were computed using Newton-Raphson load flow program in Matlab.

Rajam *et al*., discussed how network reconfiguration problem has been solved for load balancing, power loss reduction, improving operational conditions, etc. by using various EAs such as BPSO and neuro-fuzzy techniques [22]. Network reconfiguration is a complex combinatorial constrained non-linear optimization problem that is solved in a heuristic manner to achieve a given objective. In this paper the researcher proposes an EA, Modified Plant Growth Simulation Algorithm (MPGSA) in network reconfiguration with and without DG. The constraints and the objective function are dealt with separately. PGSA can handle problems with load and generation variation with time since it is based on growth process of plants, the root is the initial growth taken as the initialization and the branches and tree trunk growth at the nodes will be taken as the search for an optimal solution. With the modification made it will converge faster. The sum of power loss after reconfiguration with DG connected is taken as the objective function to maximize on the total loss reduction subject to current, power flow and voltage constraints given below:

$$Maxf = \max.(\Delta P_{Loss}^{R} + \Delta P_{Loss}^{DG}) \tag{9}$$

Subjected to $V_{\min} \le |V_K| \le V_{\max}$ and $|I_{k,k+1}| \le |I_{k,k+1,\max}|$

$$\sum_{k=1}^{n} P_{GK} \le \sum_{k=1}^{n} \left(P_K + P_{Loss,k} \right) \tag{10}$$

Where

ΔP_{Loss}^{R} -Total power loss reduction due to reconfiguration.

ΔP_{Loss}^{DG} - Total power loss reduction due to DG connection.

$V_{\min}$ -minimum bus voltage; $V_{\max}$ -maximum bus voltage.

$I_{k,k+1}$ -current through the branch between *k* and *k+1* bus.

$I_{k,k+1,\max}$ - maximum current through the branch between *k* and (*k+1)* bus.

Since the main contribution is in fixing a value for β as 0.5 (a selected random number, which enable prioritization in the growth process), thus the algorithm is called modified PGSA. It is based on number of trial and error combination values tried between 0 to 1 disabling any random search for β value making the algorithm faster and more suitable for real-time applications. The proposed method was implemented using Matlab program on an IEEE 33-bus radial distribution system with simultaneous reconfiguration and DG placement using the modified PGSA, which gave better results and faster convergence compared to ordinary PGSA. MPGSA also performed better than other EAs such as GA, refinery GA, ACO in hypercube framework and Harmony Search Algorithm (HAS), making it more suitable for practical applications on a real distribution system.

Bradan *et al.*, [23] proposed different metaheuristic algorithms in solving a multi-objective network reconfiguration problem with optimal DG output. The main objective in this paper is to simultaneously maximize the DG output, minimize the real power losses and voltage profile improvement. The objective function given in the equation (11) below represents the total fitness:

$$F = w1 \times \left(P_{loss}^{R} + IVD \right) + w2 \times \left(\frac{1}{DG_{output}} \right) \quad (11)$$

Where

P_{loss}^{R} -net power loss to be minimized,

IVD - voltage profile index to be minimized,

DG_{output} -distributed generation output to be maximized.

The multi-objective problem was solved using different optimization techniques for validation purposes. Some of the methods used were PSO, EP, (Firefly Algorithm) FA and Gravitational Search Algorithm (GSA). However, only the GSA technique implementation is described since the other techniques have already been discussed. GSA is based on the law of gravity interaction with mass of objects. The force of gravity pulls all objects towards a heavier mass. The masses will represent the population of individuals, measured by the masses' position to find their performances. A solution represented by each position while the fitness function is given by the gravitational and inertia masses. According to the law of attraction, the object with the heavier mass will attract all other objects and, in this case, heavy masses are a sign of good solutions and move slower towards lighter masses (bad solutions).

The proposed method was tested on an IEEE 33-bus distribution system in Matlab program. The optimal solutions were obtained for the tie switches and the real power loss out of the DG, which were determined simultaneously. Three cases were analyzed to evaluate on the validity of the proposed strategy. The first case aimed to minimize the power losses by simultaneous network configuration with optimal DG output. FA produced better results compared to the other techniques with losses as minimum as 72.436kW which is about 64.25% power loss reduction. The second case looked at minimizing the power losses and improving the voltage profile index by simultaneous network reconfiguration with optimal DG output and FA still out-performed the other techniques. The third case minimized power losses, improved voltage profile index and maximized the DG output by simultaneous network reconfiguration with optimal DG output. FA still emerged superior over EP, PSO and GSA. Case 2 had lesser power losses compared to case 3 hence optimizing the DG output lead to more power losses than maximizing it. A high quality and robustness in realizing an optimal network reconfiguration and DG output was presented in the proposed approach. The results also showed that the voltage profile index performed better when improved with minimizing power losses than in other cases.

In [24], Nayak discusses on how loss sensitivity analysis is used to identify optimal location for DG units in feeder reconfiguration problem solved by an EA called Hyper-Cube framework Ant Colony Optimization (HC-ACO) algorithm. The main characteristics of the proposed algorithm are as follows:

- Positive feedback - rapid search for a global solution.
- Distributed computation – to avoid premature convergence.
- Constructive greedy – heuristics assist in finding quicker acceptable solutions.

ACO algorithm is based on the colony of ants in communication when solving a problem and adapting to a situation with heuristic information and acquiring search experience in the process. The model of selection of a proper algorithm depends on the application. The HC framework will provide the automatic scaling of the auxiliary fitness function in the search process based on changing the pheromone update rules used in ACO algorithm. This will enable a robust and easier implementation of an ACO procedure.

The load flow analysis is performed using the backward-forward sweep method due to its computational efficiency and robustness in convergence making it suitable for distribution load flow analysis. The objective function is to minimize the real power losses and improve the voltage profile with given constraints. Artificial ants' movement through the buses will also help in selecting sectionalizing switches to minimize total power losses in the system. The proposed algorithm is validated in the IEEE 33 bus radial distribution system implemented using Matlab program. The presence of DG units in network reconfiguration problem had effective performance of the overall system. Some of the merits of the algorithm are fewer iterations to reach optimum solution and fast optimum convergence.

In [25], the application of GA with variable population size (GAVAPS) is proposed in solving reconfiguration problem. If the population is allowed to adapt in size according to the status of the GA search, a more efficient solution can be obtained compared to the standard GA (SGA). A population that evolves through successive generations will be used in the reconfiguration problem with the help of the proposed algorithm in searching for a solution. A possible solution is represented in the population by each individual (chromosome). Locations of open switches in the power

distribution network will be contained as a string coded by each chromosome with a lifetime parameter assigned at its creation. Each generation, creates an offspring population from the current population using the GA operators. If the chromosome exceeds the lifetime parameter, they will die and leave the genetic pool. The proposed algorithm was tested on the IEEE 33 bus radial distribution system and the performance was compared to the SGA. Significant decrease in computational cost and better exploration of the solution space was noticed with the GAVAPS algorithm.

V. DISCUSSION

In this review, it has been seen that network reconfiguration in the distribution system plays a major role in network planning and operational management in order to come up with an efficient, reliable and cost-effective system. Normally, the researchers had the same basic objective to minimize power losses despite different approaches in methodology and implementation. Other additional objectives, which have been used include cost minimization of power generated, improving on switch time, power quality improvement, improving the bus voltage deviation, minimizing cost of DG integration and balancing the load at the feeder as well as doing away with overload conditions.

Table I: Simulation results for different algorithms for IEEE 33-bus distribution system

Method	Tie-switches	Power Losses (kW)	Min. Voltage (pu)
GA [18]	7, 9, 14, 32, 37	139.54	0.9378
MCPSO[11]	7, 9, 14, 32, 37	138.93	0.9423
HC-ACO[24]	7, 14, 9, 32, 37	136.30	0.9385
MPSO [19]	7, 9, 14, 32, 37	136.36	0.9400
EPSO[21]	11, 28, 32, 33,34	120.7	0.9980
HGAPSO[17]	10, 7, 14, 32, 27	140	0.9423
GAVAPS[25]	7, 12, 31, 35,37	114.27	-
CSA [1]	7, 9, 14, 32, 37	139.51	0.9378
PSO[19]	33, 28, 34, 8, 17	149.8	0.9310

At the end of reconfiguration, power supply must be ensured to reach all the loads and the radial structure must be maintained. Optimal sizing of the DG in the distribution network also gave a high reduction in power losses. When performing reconfiguration, DG placement and sizing should not be neglected because it is also a major factor in reducing the system losses. The combination of the two techniques gave a tremendous improvement in loss reduction especially when the two are performed simultaneously together. When DG units are connected, there will be a significant improvement in voltage profile. The opening of tie and sectionalizing switches during reconfiguration lead to voltage transfer from one feeder to another and balancing it with the reduction in real power losses leading to a voltage profile improvement.

The methods applied in reconfiguring the network utilized various objectives and limited constants based on the algorithm. Heuristic methods tend to be very fast, but do not reach global solutions while meta-heuristic algorithms could reach a global solution but they have a greater computational time due to the random selections and probabilistic nature. Some of the methods require a minimum number of iterations. Table I shows the effectiveness in optimal reconfiguration of the network with the different algorithms. Table II shows some of the merits and drawbacks of some of the algorithms discussed in this work.

VI. RECOMMENDATIONS

Most research has been focused on network reconfiguration with DG integration. There is still more research to be looked at with network operations in the distribution system such as VAR compensation, restructuring of transformer, conduction size and type of feeder which should be taken into consideration when performing network reconfiguration. Some recommendations are as follows:

- Taking into account dynamic loads and DG operation modes.
- Enhancement in stability of microgrid operations.
- Taking into account the transformer life loss cost.
- Improvement of algorithm suitable for real-time network reconfiguration.

Table II: Merits and drawbacks of some of the algorithms

Algorithm	Merits	Drawbacks
GA [18][9]	Easy, simple to implement, efficient to search large solution space without getting trapped in local minimum, reasonable computational time and robust method for seeking global minimum	Slow and could not find the optimal solution easily
ACO[24]	Efficient and easy to understand and code, fewer iterations to reach optimum solution and fast optimum convergence.	High iteration is needed to find the optimum solution and takes longer computational time
EP[21]	Simple and direct,	It takes time to converge, very little literature on this algorithm
CSA [1]	Simple implementation procedure and few control parameters, finds the global optimum solution fast, can deal with multi-criteria optimization problems, can be hybridized with other algorithms	Long computational time
PSO[19]	Simple and easy to implement, powerful algorithm to aid and speed up decision making, able to escape local optimal solution and can often find good solutions for complicated problems	Not designed for discrete functions optimization, and has less optimal solutions making it difficult to obtain the global optimum solution, slow convergence

- Enhance voltage stability and maximize the use of DG.
- Applying smart systems and reducing on overall pollution.
- Minimizing on generation cost by incorporating different renewable energy sources.

VII. CONCLUSION

Different optimization techniques were studied for network reconfiguration problem, mostly to reduce the power losses and improve the profile of the voltage. The results show that reduction in power losses also lead to an improved system voltage profile. Furthermore, DG integration to the network gave significant power losses especially when a simultaneous approach is applied with network reconfiguration making it more efficient in solving the distribution system problems. Each algorithm used demonstrated the optimization problem in its own unique manner in solving the network reconfiguration problem. However, some algorithms proved to be more superior when solving the same optimization problem.

APPENDIX

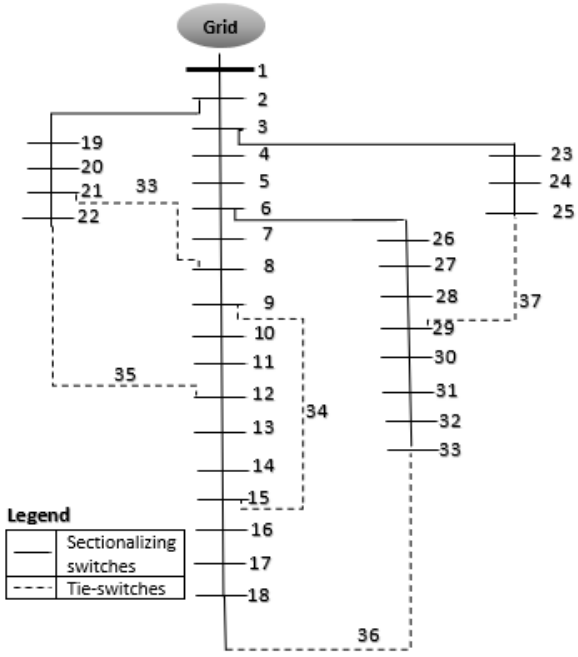


Fig. 4: IEEE 33-bus radial distribution system

ACKNOWLEDGEMENT

This research is funded by Pan African University under African Union research grant.

REFERENCES

[1] T. T. Nguyen and A. V. Truong, "Distribution network reconfiguration for power loss minimization and voltage profile improvement using cuckoo search algorithm," *Int. J. Electr. Power Energy Syst.*, vol. 68, pp. 233–242, 2015.

[2] A. Alam and F. Khatod, "Reconfiguration of Radial Distribution Network for Loss Reduction with Distributed Generation," pp. 161–164, 2011.

[3] C. Su and C. Lee, "Feeder reconfiguration and capacitor setting for loss reduction of distribution systems," vol. 58, pp. 97–102, 2001.

[4] T. Thakur, "Study and Characterization of Power Distribution Network Reconfiguration," vol. 0, pp. 1–6, 2006.

[5] N. Gupta, A. Swarnkar, and K. R. Niazi, "Distribution network reconfiguration for power quality and reliability improvement using Genetic Algorithms," *Int. J. Electr. Power Energy Syst.*, vol. 54, pp. 664–671, 2014.

[6] N. Gupta, A. Swarnkar, and K. R. Niazi, "Reconfiguration of distribution systems for real power loss minimization using adaptive particle swarm optimization," *Electr. Power Components Syst.*, vol. 39, no. 4, pp. 317–330, 2011.

[7] J. M. Rupa and S. Ganesh, "Power Flow Analysis for Radial Distribution System Using Backward / Forward Sweep Method," *Int. J. Electr. Comput. Energ. Electron. Commun. Eng.*, vol. 8, no. 10, pp. 1537–1541, 2014.

[8] A. Coelho, A. B. Rodrigues, and M. G. Da Silva, "Reliability Constraints," no. November, pp. 21–24, 2004.

[9] O. Badran, S. Mekhilef, H. Mokhlis, and W. Dahalan, "Optimal reconfiguration of distribution system connected with distributed generations : A review of different methodologies," vol. 73, no. February, pp. 854–867, 2017.

[10] S. P. Karthikeyan, K. S. Kumar, A. S. Harissh, I. J. Raglend, and D. P. Kothari, "A Matlab / GUI based simulation tool to solve load flow program for standard test systems," no. Seiscon, pp. 456–461, 2011.

[11] S. Jena, S. Jena, and S. Chauhan, "Solving distribution feeder reconfiguration and concurrent dg installation problems for power loss minimization by multi-Solving Distribution Feeder Reconfiguration and concurrent DG Installation problems for Power loss Minimization by Multi Swarm Coo," no. January, 2018.

[12] J. Zhu, "Optimal Reconfiguration Of Electrical Distribution Network," in *Optimization of Power System Operation*, Second., IEEE Press, Ed. New Jersey: John Wiley & Sons, 2015, pp. 417–420.

[13] S. Mohammad-Azari, O. Bozorg-haddad, and X. Chu, *Studies in Computational Intelligence 720 Advanced Optimization by Nature-Inspired Algorithms*. Singapore: Springer Nature Singapore, 2018.

[14] J. Naish, "Managing chaos.," *Nurs. Manag. (Harrow).*, vol. 2, no. 1, p. 3, 1995.

[15] C. Daniel. L. Khan and S. Ravichandran, "Reduction Using Ant Colony System Algorithm," pp. 1–4, 2005.

[16] O. Badran, S. Mekhilef, H. Mokhlis, and W. Dahalan, "Optimal switching sequence path for distribution network reconfiguration considering different types of distributed generation," *IEEJ Trans. Electr. Electron. Eng.*, vol. 12, no. 6, pp. 874–882, 2017.

[17] A. Wazir and N. Arbab, "Analysis and Optimization of IEEE 33 Bus Radial Distributed System Using Optimization Algorithm," *JETAE) J. Emerg. Trends Appl. Eng.*, vol. 1, no. 2, pp. 2518–4059, 2016.

[18] P. Rao and V. Reddy, "Impact of Distribution Feeder Reconfiguration for Loss Reduction on Bus Voltages -A Perspective," *Int. J. Adv. Res. Ideas Innov. Technol.*, vol. 3, no. 3, pp. 137–143, 2017.

[19] I. I. Atteya, H. A. Ashour, N. Farmi, and D. Strickland, "Distribution Network Reconfiguration in Smart Grid System Using Modified Particle Swarm Optimization," vol. 5, 2016.

[20] X. Jin, K. Li, and I. Introduction, "Particle Swarm Optimization," no. November, pp. 21–24, 2004.

[21] M. F. Sulaima, H. Mokhlis, and I. J. Hazriq, "A DNR Using Evolutionary PSO for Power Loss Reduction," *J. Telecommun. Electron. Comput. Eng.*, vol. 5, no. 1, pp. 31–36, 2013.

[22] R. Rajaram, K. S. Kumar, and N. Rajasekar, "Power system reconfiguration in a radial distribution network for reducing losses and to improve voltage profile using modified plant growth simulation …," *Energy Reports*, vol. 1, pp. 116–122, 2015.

[23] O. Badran, H. Mokhlis, S. Mekhilef, and W. Dahalan, "Multi-Objective Network Reconfiguration with Optimal DG Output Using Meta-Heuristic Search Algorithms," *Arab. J. Sci. Eng.*, 2017.

[24] M. R. Nayak, "Optimal feeder reconfiguration of distribution system with distributed generation units using HC-ACO," *Int. J. Electr. Eng. Informatics*, vol. 6, no. 1, pp. 107–128, 2014.

[25] M. Abdelaziz, "Distribution network reconfiguration using a genetic algorithm with varying population size," *Electr. Power Syst. Res.*, vol. 142, pp. 9–11, 2017.

Quadcopter Attitude Estimation using filters for Sensor Fusion in 6D Inertial Measurement Unit

Jackson O. Oloo and Stanley I. Kamau

***Abstract*—** Orientation tracking of a quadcopter Unmanned Aerial Vehicle (UAV) involves monitoring the Roll, Pitch and Yaw angles. These angles provide feedback information that is then used to give appropriate angling and heading orientation. Measurements of these Euler angles is accomplished by use of an Inertial Measurement Unit (IMU) consisting of either a gyroscope, accelerometer or both. The IMU created with the gyroscope is less sensitive to vibrations and is not affected by earth's gravity. One of the problems that a gyro based IMU encounters is the drifting of the angles. Another problem occurs when the IMU is started at an angled surface. This is because the IMU has no reference to what is level. In static or slow movement, the accelerometer measures roll and pitch by leveling to correct the gyro-unbounded error. This is due to the trustworthiness of the gravitational measurement. While the accelerometer gives absolute measurement of the quadcopter attitude, the motors on the quadcopter produce a lot of vibrations introducing significant noise into the accelerometer reading. Therefore, a proper fusion of IMU data is needed to overcome the shortcomings of each sensor. Kalman filter is therefore proposed to merge the two sensor measurements to achieve better estimates, redundancy and drift compensation. In conclusion, the performance of the Kalman filter is then compared with that of the unfiltered sensor data and Complimentary filter.

***Keywords*—** Accelerometer, Complimentary filter, Gyroscope, Kalman filter,

I. Introduction

A quadrotor is a helicopter lifted and propelled by four rotors. Small sized quadrotors are often used as Unmanned Aerial Vehicles (UAVs) in research and amateur projects, because of the simple symmetric structure and relatively easy control law with respect to traditional helicopters.

Quadrotors have a set of sensors that provide the information needed by the attitude, altitude and the navigation control systems. This set of sensors is usually called an Inertial Measurement Unit (IMU). The IMU of a quadrotor contains the following sensors: an accelerometer, a gyroscope, a magnetometer and a barometer. During flight, the motors in the quadrotor introduces noise into the data read from the sensors. This brings divergence from the intended orientation and trajectory.

Filtering involves seeking for the best values of the system states via new measurements and updating of the new measurements [1]. Kalman filter has become popular and is used in almost every sensor processing applications. Some extensions of Kalman filter are adaptive Kalman filtering [2], unscented Kalman filtering [3] ,extended Kalman filtering [4]. The most and widely used filtering techniques are based on extended Kalman filter [5], [6]. Some other techniques have also been developed like the nonlinear observer given in [7], or based on unscented filter [8]. Most of these methods are computationally demanding. Estimation of Quadrotor attitude using Extended Kalman Filter (EKF) has been discussed in [9] while [10] compares the performance of the quadrotor with both the EKF and the Kalman Filter with the conclusion that EKF gives the best performance theoretically. However, attitude estimation using EKF has been found inapplicable to embedded systems.

The use of Kalman filters to estimate the attitude continues to attract many researchers. In [11], Kalman filter is designed to estimate the noisy states of the system. However, it is computationally demanding and difficult to understand. An alternative, the Complementary filters, which are not so computationally demanding are used for attitude estimation in [12], [13] and its performance is compared with that of Kalman filter in [14]. Accelerometer and gyroscope measurements are fused using Complimentary filter in [15] to estimate the orientation.

In this scheme, the Kalman filter and complementary filter is applied to estimate the attitude states of the quadrotor from the noisy measurements of on board Microelectromechanical sensors (MEMS). The estimated state is intended to be used by a control algorithm (not discussed in this work) to maintain the desired attitude during various maneuvers. In conclusion, the performance of the Kalman filter is then compared with that of the unfiltered sensor data and Complimentary filter.

II. MicroElectroMechanical Sensor (MEMs)

A. Accelerometer Model

Accelerometer measures total acceleration relative to free fall, also called specific force $\bar{f}^b$ [16]
However, Accelerometer do not capture the high frequency dynamics. When an accelerometer is part of a moving system like UAVs and robots, it not only measures acceleration due to gravity but also translational and rotational accelerations. Therefore, an ideal accelerometer aligned with the body measures specific force as shown in (1). A detailed derivation is given in [17].

J. O. Oloo, Department of Electrical and Electronic Engineering, JKUAT (phone: +254727422602; e-mail: joloo@jkuat.ac.ke).

S. I. Kamau, Department of Electrical and Electronic Engineering, JKUAT (e-mail: skamau@eng. jkuat.ac.ke).

$$\overline{f}^{b} = \begin{bmatrix} f_{x,accel} \\ f_{y,accel} \\ f_{z,accel} \end{bmatrix} = -g \begin{bmatrix} -\sin\theta \\ \sin\phi\cos\theta \\ \cos\phi\cos\theta \end{bmatrix} \quad (1)$$

Where, $\bar{g}^{b}$ is gravity in body coordinates, $\emptyset$ and θ represent roll and pitch in radian respectively.

B. Gyroscope model

Gyro sensors measure angular velocity in x, y, z directions although its measurements include biases. It is modeled as follows:

$$\overline{\Omega}^{b}_{gyro} = \begin{bmatrix} \Omega_x \\ \Omega_y \\ \Omega_z \end{bmatrix} = \overline{\Omega}^{b} + \begin{bmatrix} n_{\Omega_x} + b_{\Omega_x} \\ n_{\Omega_y} + b_{\Omega_y} \\ n_{\Omega_z} + b_{\Omega_z} \end{bmatrix} \quad (2)$$

Where, b_{Ω} and n_{Ω} represent the gyro bias and the associated noise respectively. Gyro measurement and Euler angle rate are related as shown [17]:

$$\begin{bmatrix} \dot{\phi} \\ \dot{\theta} \\ \dot{\varphi} \end{bmatrix} = \begin{bmatrix} 1 & \sin\phi\tan\theta & \cos\phi\tan\theta \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi\sec\theta & \cos\phi\sec\theta \end{bmatrix} \begin{bmatrix} \Omega_x \\ \Omega_y \\ \Omega_z \end{bmatrix} \quad (3)$$

Where, $\emptyset$ and θ represent roll and pitch in radian respectively and Ω is the propeller angular velocity.

Errors accumulate with time due to gyro bias making it practically impossible to rely on gyro data alone for Euler angle estimation. Hence, accelerometers are used to compensate for the gyro's drifts in pitch and roll estimation. (Yaw estimation is not covered in this work).

III. Mathematical Formulation of Kalman Filter

Kalman filter is a recursive filter that estimates the states of the dynamics of a system by noisy measurement.

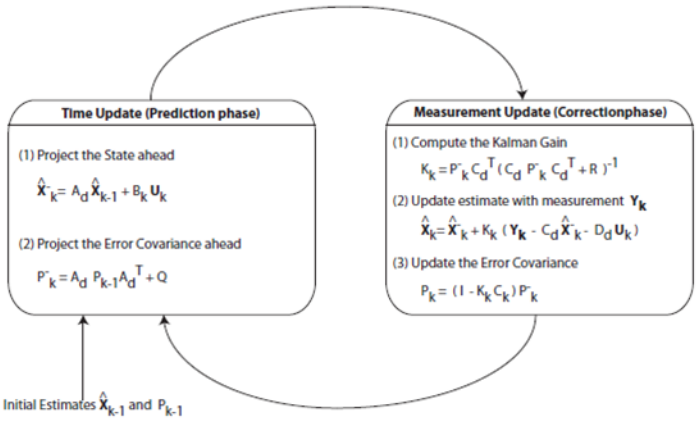


Fig. 1 Kalman model

The Kalman filter is based on a two-step process: First, the system acts as a Predictor; i.e. it uses the model of the system, the current state and the input vector to predict the future state considering the covariance error. In application, the filter takes the gyroscope measurements and calculates attitude estimations based on the gyroscope rates, and makes a prediction estimate of the error covariance.

The Measurement update phase, which is the second phase, corrects the predicted state and the estimated covariance error according to the measurements and its noise covariance. These are then used to calculate the Kalman gain. The accelerometer data is incorporated to aid the gyroscope measurement. These two values are multiplied by the Kalman gain to use a percentage of each measurement based on their noise characteristics.

Therefore, a model for the prediction of the angular velocity (without the model noise) is given by [18]:

$$\dot{\theta}_k = \dot{\theta}_{k-1} + \frac{k}{j}(u_1 - u_2) \quad (4)$$

Where $\dot{\theta}_k$ is the angular velocity, k is a constant for the linear relationship between the force generated by the motor and the input, j is the inertia and u is the input to the motors. To make (4) complete, the model noise $w_k \sim N(0, Q)$ is introduced to take model errors into consideration. Q is the covariance matrix of the noise given. The predicted state is then updated according to the steps described in [18].

IV. Complimentary Filter

When measuring the body angle with the accelerometer, it is affected by translation and vibrations of the motors, but the errors are not accumulated. When measuring with the gyro sensor, the errors are accumulated, but vibrations do not affect its operation. These two sensors measure the same physical quantities, and the properties are complementary, so the

weaknesses of each sensor can be supplemented through convergence.

Generally, all the forces working on the object are measured by accelerometer and as the small forces creates disturbance in measurement, long-term measurement is reliable. So for accelerometer low pass filter is needed for correction. In the gyroscopic sensor the integration is done over a period of time and the value starts to drift in the long term, so high pass filter is needed for gyroscopic data correction[19], [20]. Therefore, the complementary filter consists of both low and high pass filter as shown in Fig. 2.

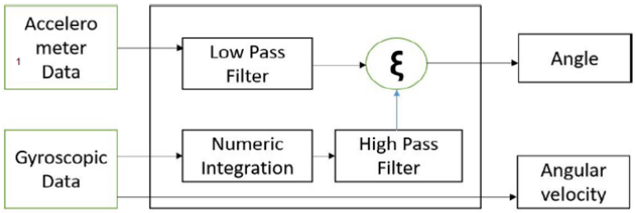


Fig. 2 Complimentary filter block diagram

The complimentary filter is a unity filter i.e. $Gain1(s) + Gain2(s) = 1$ and is based on time constant to produce desired gains [21]

$$\alpha = \frac{\tau}{\tau + dt} \tag{5}$$

τ is calculated from examining the gyroscope drift rate.

The complimentary filter angle is therefore calculated by summing weighted portions of the gyroscope and accelerometer angles to create a more accurate combined attitude angle.

$$\theta_k = gyro\_gain * (\theta_{k-1} + gyro * dt) + (0.020) * (accel) \tag{6}$$

V. Implementation

In this scheme, a 6-degree of freedom (6-DOF) MEMS sensor MPU-6050 has been used. It combines a 3-axis gyroscope, 3-axis accelerometer, and a Digital Motion Processor™ (DMP) all in a small 4x4x0.9mm package.

The fusion algorithms were implemented in Arduino platform.

The test was performed as follows:

- First the IMU GY-521breakoutboard was tilted smoothly.
- Next, the board was then continually tilted with some vibrations, i.e. by tapping and shaking the board quickly

The data was received on the Arduino Uno serial monitor. A program was written on the Arduino environment to prompt for inputs from the IMU sensor. The received inputs are then processed using the Kalman and Complimentary libraries.

Then from the data, their performance was analyzed using Matlab.

VI. Simulation Results

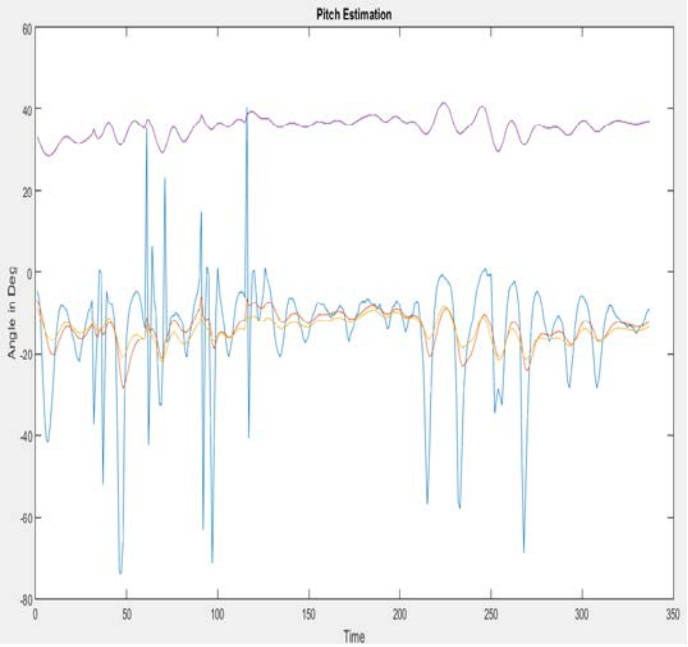


Fig. 3 Pitch angle estimation using Kalman Filter, Complimentary filter and MEMs sensors raw data

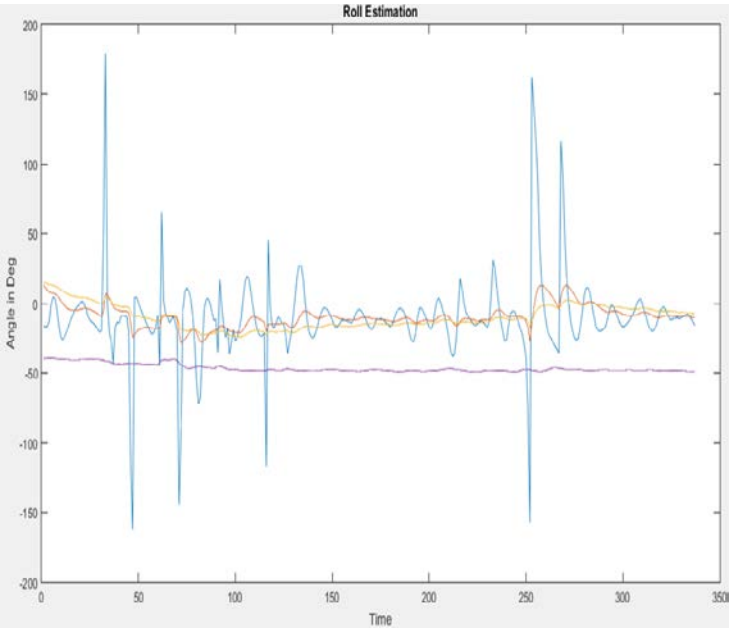


Fig. 4 Raw angle estimation using Kalman Filter, Complimentary filter and MEMs sensors raw data

From the Figs. 3 and 4, gyro data is represented by the purple line, Accelerometer by the blue line, the yellow line is the filtered data by Kalman filter and the red line is the complimentary filtered data. The filtered signal was obtained by combining the Accelerometer and Gyroscope data using the two fusion algorithms.

The purple line clearly shows how the gyro data drifts slowly with time from the zero level while the blue line shows the effect of vibrations (shaking) on the accelerometer data. The

gyro data is however not affected by vibrations while the accelerometer data has no bias.

It can be observed that even with much vibration, the Kalman and Complimentary Filtered data is not much affected and the noise pronounced in the MEMs sensor raw data is eliminated with these two filters.

VII. CONCLUSION

From the results, the Kalman filter is more precise than the Complementary Filter, especially during vibrations. The Kalman filter is however mathematically involving and difficult to understand as compared to the complimentary filter which is easy to implement. Future work would involve application of Extended Kalman filter to perform the fusion of the two sensor signals and compare its performance with that of linear Kalman Filter studied in this work and to implement these Filter attributes in a physical Quadcopter system.

REFERENCES

[1] Gibbs B., “Advanced Kalman Filtering, Least Squares and Modelling”, 2011, N.Y, Wiley and Sons Publishers. Chapter 8.

[2] W.Li and J.Wang, "Effective Adaptive Kalman Filter for MEMSIMU/ Magnetometers Integrated Attitude and Heading Reference Systems”, the Journal of Navigation, 2012, p: 99 – 113.

[3] M. Rhudy.,Y.Gu ,J.Gross J., "Evaluation of Matrix Square Root Operations for UKF within a GPS/INS Sensor Fusion Application", International Journal of Navigation and Observation, 2011.

[4] X.Chen, Y.Xu, Q.Li, "Application of Adaptive Extended Kalman Smoothing on INS/WSN Integration System for Mobile Robot Indoors", Mathematical Problems in Engineering, 2013.

[5] J. Crassidis, F. Markley, and F. Cheng, “Survey of nonlinear attitude estimation methods,” Journal of guidance, control, and dynamics, vol. 30, no. 01, pp. 12–28, January 2007.

[6] M. Jun, S. Roumeliotis, and G. Sukhatme, “State estimation of an autonomous helicopter using kalman filtering,” in Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 3, 1999, pp. 1346 – 1353.

[7] A. El Hadri and A. Benallegue, “Attitude estimation with gyros-bias compensation using low-cost sensors,” in Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference, Shanghai, P.R. China, December 16-18 2009, pp. 8077–8082.

[8] J. Crassidis and M. F.L., “Unscented filtering for spacecraft attitude estimation,” Journal of guidance, control, and dynamics, vol. 26, no. 4, pp. 536–542, 2003

[9] L. Ascorti, “An application of the extended Kalman filter to the attitude control of a quadrotor”, Mastersthesis, PolitecnicodiMilano,2012.

[10] Wenjing Liang, “Attitude Estimation of Quadcopter through Extended Kalman Filter”, Masters Thesis, Lehigh University ,2017.

[11] H.G. de Marina and F. J. Pereda, “UAV Attitude Estimation Using Unscented Kalman Filter and TRIAD”, IEEE Transaction on Industrial Electronics, vol. 59, no. 11, 2012.

[12] M. Euston, P. Coote, R. Mahony, J. Kim, and T. Hamel, “A complementary filter for attitude estimation of a fixed-wing uav,” in IEEE/RSJ International Conference on Intelligent Robots and Systems, Acropolis Convention Center, Nice, France, Sept, 22-26 2008, pp. 340–345.

[13] J. Vasconcelos, C. Silvestre, P. Oliveira, P. Batista, and C. B., “Discrete time-varying attitude complementary filter,” in American Control Conference Hyatt Regency Riverfront, St. Louis, MO, USA, June 10-12 2009, pp. 4056–4061.

[14] J. W. T. Higgins, “A comparison of complementary and kalman filtering,” IEEE Transaction On Aerospace And Electronic Sysytems, vol. AES-1 1, no. 3, pp. 321–325, May 1975.

[15] V. Kubelka and M. Reinstein, “Complementary filtering approach to orientation estimation using inertial sensors only,” in IEEE International Conference on Robotics and Automation, RiverCentre, Saint Paul, Minnesota, USA, May 14-18 2012, pp. 599–605.

[16] D. D. Quoc, “Attitude Estimation Algorithms Using Low Cost IMU”, International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 8, No. 2 (2015), pp. 313-324.

[17] A.Kaba, A. Enneydan and Kiyak, “ Model Derivation, Attitude Control and Kalman Filter Estimation of a Quadcopter”, 4th International Conference on Electrical and Electronics Engineering 2017.

[18] M.Boutayeb, H. Rafaralahy, and M. Darouach. "Convergence analysis of the extended Kalman filter used as an observer for nonlinear deterministic discrete-time systems." IEEE transactions on automatic control 42.4 (1997): 581-586.

[19] D. Cao, Q. Qu, C. T. Li and L. C. He, “Research of Attitude Estimation of UAV Based on Information Fusion of Complementary Filter”, Fourth International Conference on Computer Sciences and Convergence Information Technology, IEEE Computer Society, 2009. pp. 1290-1293.

[20] Walter T. Higgins, “A Comparison of Complementary and Kalman Filtering”, 2015 10th Asian Control Conference (ASCC), pp.1-6, May-June, 2015.

[21] J. Vasconcelos, C. Silvestre, P. Oliveira, P. Batista, and C. B., “Discrete time-varying attitude complementary filter,” in American Control Conference Hyatt Regency Riverfront, St. Louis, MO, USA, June 10-12 2009, pp. 4056–4061.

Quadcopter Control Algorithms in the event of Loss of One of the Actuators: A Review

Jackson O. Oloo and Stanley I. Kamau

***Abstract*—**There are increasing needs to develop reliable control systems that can be used to automatically stabilize a Quadcopter Unmanned Aerial Vehicle (UAV) when one of the four rotors is under fault. A fault tolerant mechanism that extends the capabilities of the quadcopter system to operate under the presence of faults is of interest to the research community. The quadcopter UAV is a great platform for control systems research as its nonlinear nature and under-actuated configuration makes it ideal to analyze control algorithms. The nonlinearity is even further pronounced when there is an actuator fault. This review gives an overview of Quadcopter UAV dynamic system. Considering a single actuator fault scenario, applicable algorithms are analyzed highlighting their advantages and disadvantages. The algorithms include PID control, Gain-Scheduling PID, Linear Quadratic control, Backstepping, Feedback linearization, Sliding Mode control, Model Predictive control and Model Reference Adaptive Controller. The conclusion of this work is a proposal of combination of positive attributes of each algorithm to compensate for the limitations of the other.

***Keywords*—** Control Algorithms, Nonlinear Control, Quadcopter

I. Introduction

A quadcopter is an aircraft that is lifted and propelled by four rotors in a cross configuration and its basic motions are generated by varying the speeds of all the four rotors. It is a 6 Degree of Freedom (DOF) device with only four actuators, which makes it an under actuated vehicle with unstable dynamics. For small aerial vehicles, due to hardware redundancy limitations, design of a reliable control system plays an important role in ensuring acceptable and efficient performance.

In recent years, there has been a surge of interest in the use of small Unmanned Aerial Vehicles (UAVs) for various civilian and military applications [1]. These applications include package delivery, aerial imagery, surveillance, and structural inspection; a common aspect is that these tasks are either in remotely inaccessible locations and require dangerous maneuverability or are in unfriendly environments in case of military operations. Several different UAV platforms exist that have the potential to solve these problems such as fixed-wing airplanes, lighter-than-air blimps, and multirotor aircrafts. A quadcopter has advantages over the fixed wing UAVs in that it has Vertical Take-off and Landing (VTOL) capabilities and can perform maneuvers. Its advantage over other rotary UAVs, such as a helicopter, is that it is mechanically simple; a quadcopter does not need a complex set of mechanical linkages to alter rotor blade angles. Quadcopter helicopters do not require a tail rotor and this allows it to devote all vehicle power o producing lift. This allows for significant payload capacity in relation to vehicle weight.

However, a quadcopter is a six degrees of freedom system with only four actuators, making it underactuated as well as being a highly nonlinear and unstable system [2]. With such a configuration, the entire vehicle must tip in one direction or another in order to direct the rotor thrusts to actuate lateral or longitudinal motion. This could be seen as a potential disadvantage as it does constrain the dynamics of the vehicle in that it cannot cause acceleration forward or back or from side to side while maintaining a given orientation.

It is therefore a bigger challenge to maintain full control of all the attitude states and all the translational states when one of the rotors has failed and the system becomes even further under actuated. This makes the quadcopter highly non-linear and several uncertainties are encountered during its missions. This has led to several control algorithms proposed in the literature. In this work, a review of the prominent controllers applied to the quadrotor is reviewed.

II. QUADCOPTER MODEL

A. MOTION OVERVIEW

Quadcopters consist of four rotors attached to a rigid cross airframe as shown in Fig.1, with two opposing rotors rotating clockwise (1,3) and the other two rotating counterclockwise (2,4). Control of quadcopter is achieved by differential control of the thrust generated by each rotor.

Attitude motion is accomplished by simultaneously increasing or decreasing the speed of all four rotors. Pitch angle,θ, is controlled by speeding up motor 3 while slowing down motor 1 or vice versa while roll angle, φ, is controlled by slowing down motor 4 while speeding up motor 2 or vice versa. Then yaw angle,ψ, is controlled by speeding up motors 1 and 3 while slowing down motors 2 and 4 or vice versa.

J. O. Oloo, Department of Electrical and Electronic Engineering, JKUAT (phone: +254727422602; e-mail: joloo@jkuat.ac.ke).

S. I. Kamau, Department of Electrical and Electronic Engineering, JKUAT (e-mail: skamau@eng. jkuat.ac.ke).

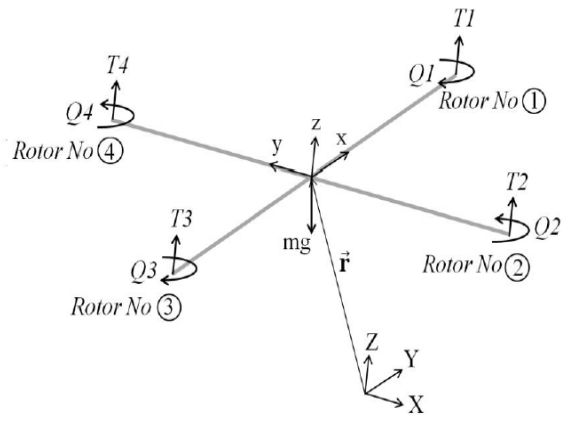


Fig 1 Quadrotor free body diagram.

B. EQUATIONS OF MOTION

The equations of motion that describe the dynamics of the vehicle are developed using a Newton-Euler formalism. The system in state space form is given as:

$$\dot{X} = f(x, U) \tag{1}$$

Where U is the input vector and x the state vector.

The derivations of the nonlinear dynamics of the system is summarised by the following equations [3] [4].

$$\dot{X} = \begin{bmatrix} \dot{\phi} \\ \ddot{\phi} \\ \dot{\theta} \\ \ddot{\theta} \\ \dot{\varphi} \\ \ddot{\varphi} \\ \dot{z} \\ \ddot{z} \\ \dot{x} \\ \ddot{x} \\ \dot{y} \\ \ddot{y} \end{bmatrix} = \begin{bmatrix} \dot{\phi} \\ \dot{\theta}\dot{\varphi}\frac{I_{YY}-I_{ZZ}}{I_{XX}} + \dot{\theta}\,\frac{J_r}{I_{XX}}\Omega_r + \frac{l_a}{I_{XX}}U_2 \\ \dot{\theta} \\ \dot{\theta}\dot{\varphi}\frac{I_{ZZ}-I_{XX}}{I_{YY}} - \dot{\phi}\,\frac{J_r}{I_{YY}}\Omega_r + \frac{l_a}{I_{YY}}U_3 \\ \dot{\varphi} \\ \dot{\theta}\dot{\phi}\frac{I_{XX}-I_{YY}}{I_{ZZ}} + \frac{1}{I_{ZZ}}U_4 \\ \dot{z} \\ -g + (\cos\phi\cos\theta)\frac{U_1}{m_s} \\ \dot{x} \\ u_x\frac{U_1}{m_s} \\ \dot{y} \\ u_y\frac{U_1}{m_s} \end{bmatrix} \tag{2}$$

Where,

I_{XX}, I_{YY}, I_{ZZ} are the Moment of Inertia of the Quadcopter.

l_a Quadcopter arm length

m_s total mass of the Quadcopter

g acceleration of gravity

ϕ, θ *and* φ are the roll, pitch and yaw angles respectively

J_r moment of inertia of the rotor about its axis of rotation and

$$\begin{bmatrix} u_x \\ u_y \end{bmatrix} = \begin{bmatrix} \cos\phi sin\theta\cos\varphi + \sin\phi\sin\varphi \\ \cos\phi\sin\theta\sin\varphi - \sin\phi\cos\varphi \end{bmatrix}$$

U is the input vector consisting of U_1 (total thrust), and U_2, U_3, U_4 which are related to the rotation of the Quadcopter. The inputs are mapped by:

$$U = \begin{bmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{bmatrix} = \begin{bmatrix} b(\Omega_1^2 + \Omega_2^2 + \Omega_3^2 + \Omega_4^2) \\ b(-\Omega_2^2 + \Omega_4^2) \\ b(\Omega_1^2 - \Omega_3^2) \\ d(-\Omega_1^2 + \Omega_2^2 - \Omega_3^2 + \Omega_4^2) \end{bmatrix} \tag{3}$$

b, d thrust, drag coefficient and Ω is the angular velocity of the rotor.

A fault is generated in an actuator by multiplying its control input by a gain smaller than one, thus simulating a loss in the control effectiveness.

[4] provides a detailed quadrotor model that is derived from the Newton-Euler formulation.

III. QUADROTOR CONTROL ALGORITHMS

This section discusses both linear and nonlinear approaches towards the control of a quadcopter under a faulty rotor and compares their performance in relation to their advantages and disadvantages.

A. Proportional Integral and Derivative (PID) Technique

PID controllers are control loop feedback mechanisms that directly adjust control values with a closed-form formula based on derivative, integral, and proportional gains [5].

PID formulation is employed in [5] as shown below:

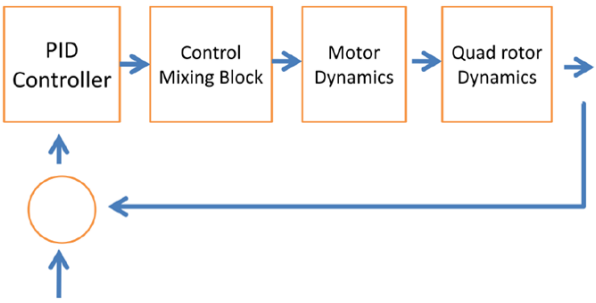


Fig 2 PID Control Strategy for a Quadrotor

Beginning with an open loop system to show the advantages of PID control, the results are as shown in Fig. 3 and 4.

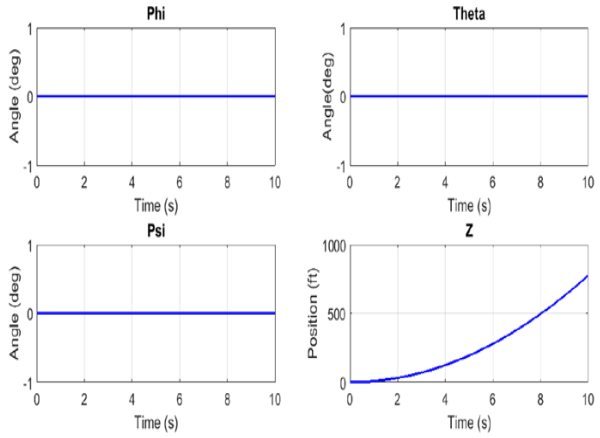


Fig. 3 Open loop step responses for motor speeds $w_1 = w_2 = w_3 = w_4 = 5000\ RPM$

Fig. 4 shows output of quadrotor response when $w_1 = w_2 = w_3 = 5000\ RPM\ \ and\ w_4 = 4950\ RPM$

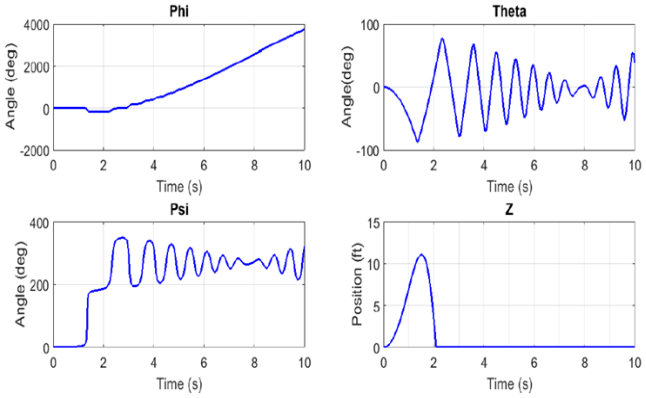


Fig. 4 Open loop step responses for motor speeds $w_1 = w_2 = w_3 = 5000\ RPM\ \ and\ w_4 = 4950\ RPM$

In [5], the scheme then describes the implementation of a PID controller without rotor failure. The gain values are set as $K_p = 10, K_i = 0$ and$K_d = 11$. The closed loop response is as shown in Fig. 6 with overshoot of 0% and settling time of 3.6043 seconds.

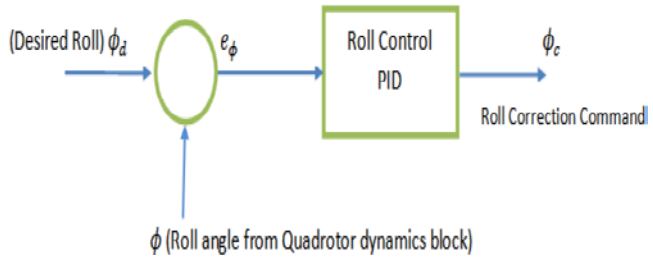


Fig. 5 Roll Control PID Block

Similar arrangement applies to PID Pitch control

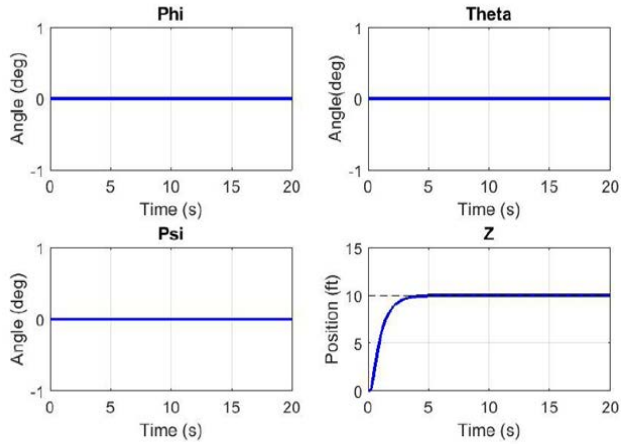


Fig. 6 Altitude Closed Loop Step Response

PID control was implemented in design of Quadrotor Controller for stabilization after failure of one of the rotors [6]. Motor 2 is switched off after 37 seconds and the response of the system is as shown in Fig. 7.

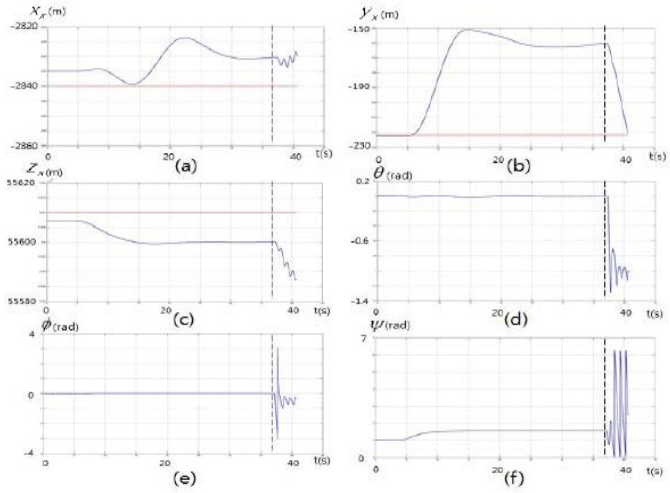


Fig. 7 Response for PID fault tolerant control

In conclusion, the classical PID controller has the advantage that parameter gains are easy to tune, simple to design and is robust. However, the quadrotor being underactuated has a nonlinear mathematical model. PID linear controller has been found to struggle with aggressive maneuvers [7] especially when one of the rotors is faulty.

B. Gain Scheduled Proportional Integral Derivative controller (GS-PID)

PID controllers are designed and tuned in both fault-free and faulty situations to control the quadcopter under normal and faulty flight conditions. In GS-PID controller, several sets of pre-tuned gains are applied to the controllers in different flight conditions under both fault-free and faulty cases. In the next step, attempts to obtain the best stability and performance of Quadcopter in attitude and altitude tracking control under both cases and to switch the controller gains from one set of pre-tuned PID controller to another set of the gains in the presence of different levels of actuator faults are carried out. One of the main parameters to be considered in GS-PID is the

switching time between the time of fault occurrence and the time of switching to new set of gains. If this transient (switching) time is held long (more than one second) it can cause the Quadcopter to hit the ground and cause a crash.

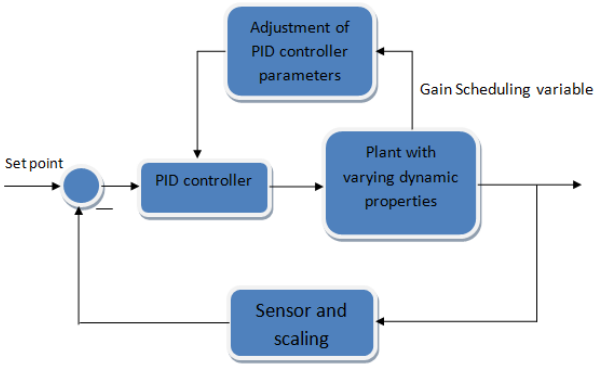


Fig. 8 GS-PID controller structures

GS-PID is applied in [8] for fault tolerant control of a quadcopter for an 18% of overall loss in power of all motors. Acceptable tracking deviation from the desired square trajectory after the fault occurrence was obtained with the fault injected at 20s.

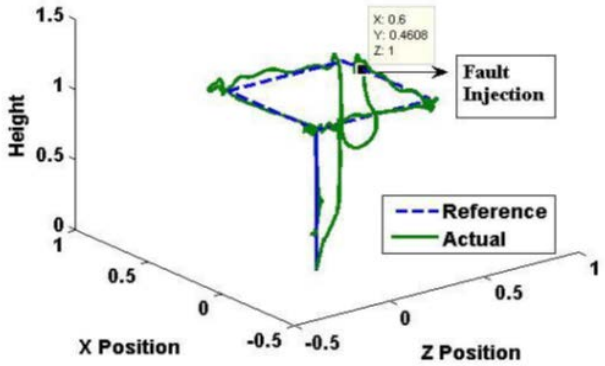


Fig. 9 GS-PID without time delay for controller switching in faulty condition

As much as GS-PID is easier to design and implement in MATLAB/ Simulink, it was observed that a good tuning for the GS-PID controller gains was very time consuming and gains could change from one flight depending on the flight environment. Any change in lab environment during flight forced the need for retuning of the gains.

C. *Model reference adaptive controller (MRAC)*

Model Reference Adaptive Control (MRAC) is concerned with forcing the dynamic response of the controlled system to asymptotically approach that of a reference system, despite parametric uncertainties (faults) in the system [9]. Two major subcategories of MRAC are those of indirect methods, in which the uncertain plant parameters are estimated and the controller redesigned online based on the estimated parameters, and direct methods, in which the tracking error is forced to zero without regard to parameter estimation accuracy.

Direct method poses an advantage over indirect method since there is no need of estimation of unknown parameters for implementing the adaptive controller. Direct method is selected in [9] for fault-tolerant control of a quadcopter.

In [9] the flight was tested for both hovering control and square trajectory tracking controls with fault injection. The experimental flight testing results are shown in Fig. 10 and 11. In Fig. 10, a fault-free condition is applied to the Quadcopter and the MRAC was able to track the trajectory close to real one. In Figure 10, a fault is injected to the left and back motors at 20 sec with a loss of 18% of power during the flight. From Fig. 11, the Quadcopter could still track the desired trajectory with a safe landing.

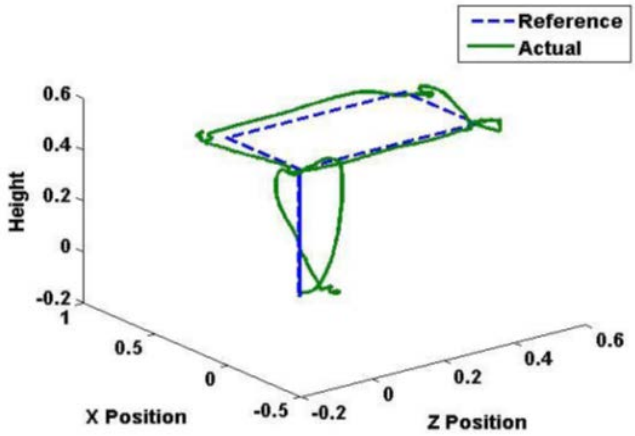


Fig. 10 Square trajectory in fault-free condition with MRAC

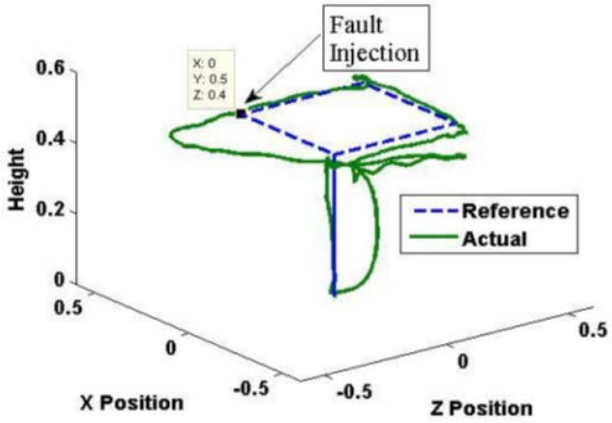


Fig. 11 Square trajectory in faulty condition (left and back motors) with MRAC

The MRAC used in [9] proved to be more reliable and robust to the lab noises and environment changes. However, involving mathematical model derivations are needed to design and implement the controller.

D. *LQR controller*

This method derives the feedback gain for a system. Applying the Linear Quadratic (LQ) control requires the system described in (1) to be linearized to $\dot{X} = AX + Bu$ i.e.

$$\dot{X}^T = [\dot{\phi}\ \ddot{\phi}\ \dot{\theta}\ \ddot{\theta}\ \dot{\varphi}\ \ddot{\varphi}]^T \quad (4)$$

A state feedback control $u = -Kx$ is designed to stabilize the system as:

$$\dot{X} = (A - BK)x \quad (5)$$

Where

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{I_{YY} - I_{ZZ}\dot{\varphi}}{2I_{XX}} & 0 & \frac{I_{YY} - I_{ZZ}\dot{\theta}}{2I_{XX}} \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{I_{ZZ} - I_{XX}\dot{\varphi}}{2I_{YY}} & 0 & 0 & 0 & \frac{I_{YY} - I_{XX}\dot{\phi}}{2I_{YY}} \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & \frac{I_{XX} - I_{YY}\dot{\theta}}{2I_{ZZ}} & 0 & \frac{I_{XX} - I_{YY}\dot{\phi}}{2I_{ZZ}} & 0 & 0 \end{bmatrix}$$

And

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{l}{I_{XX}} & 0 & 0 & \frac{Jr\dot{\theta}}{I_{XX}} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{l}{I_{YY}} & 0 & 0 & \frac{Jr\dot{\theta}}{I_{YY}} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{I_{ZZ}} & 0 \end{bmatrix}$$

The state variable feedback gain, *K* is calculated as

$$K = R^{-1}B^T P \quad (6)$$

Where *R and P* are weighted matrix and constant matrix respectively.

In [10], disturbance by means of a step signal is injected into the system and the model responds as shown in Fig. 12.

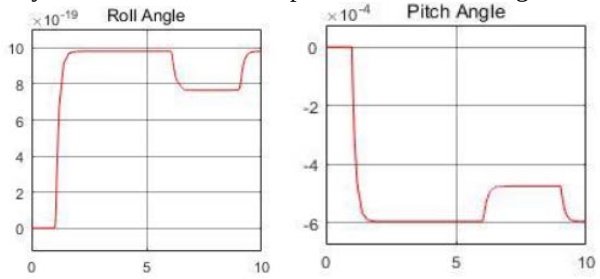


Fig. 12 Nonlinear model response for step signal disturbance

From the results, the controller tries to bring the model back to zero as shown in Fig. 12 but at the end the model does not go to zero because the given disturbance is a huge one compared to the values obtained from the model.

Castillo et al. [11] have implemented this kind of controller. During simulation, the controller has performed satisfactory. When strong perturbation was introduced, the controller due to its linearity was not able to stabilize the system. On the physical model, this controller was not able to stabilize the system at all.

However, simulation results have shown that an LQR architecture in fault free applications has superior performance compared to a classical PD architecture in terms of transient response [12]. In [13] LQR algorithm is applied to a quadrotor and its performance compared to that of the PID controller. However, the PID is applied on the quadrotor simplified dynamics and the LQR on the complete model. Both approaches provided average results but it implicitly was clear that the LQR approach had better performance considering the fact that it was applied to a more complete dynamic model.

E. Feedback Linearization (FBL)

FBL transforms a nonlinear system into an equivalent linear system through a suitable control law. A control law is designed such that the nonlinear term is cancelled to result in a controllable linear system. Some of the disadvantages of using FBL is that it needs a Fault Diagnosis and Identification system (FDIs), the loss of precision due to linearization and requiring an exact model for implementation [14].

In [15], a Fault Tolerant Controller (FTC) was developed having a double control loop architecture with an inner and outer controller. The inner and faster controller has the task to regulate the attitude angles and the altitude of the vehicle, while an outer and slower controller has the task to supply a proper input to the fault-free couple of rotors in order to make the quadrotor reach a desired position in space.

The inner control law ensures that when one of the rotor fails, the velocity of the rotor laying on the same axis of the faulty rotor is modulated until the value of the angle controlled by the faulty couple of rotors is zero. In this configuration, the quadrotor is parallel to the ground without the outer controller, spinning around the vertical axis and varying simultaneously the rotational velocity of the two rotors of the fault-free couple. This makes it possible to set a desired altitude for the vehicle.
The outer control law ensures the quadcopter reaches a desired position in space by supplying proper input to the fault free rotors.

The inner control loop controls roll, pitch and altitude, while the outer control loop sets the desired values of the $\emptyset$ and θ angles in order to control position in the $xy - plane$. Since roll, pitch and altitude have the highest priority during flight, the inner controller works much faster than the outer controller.

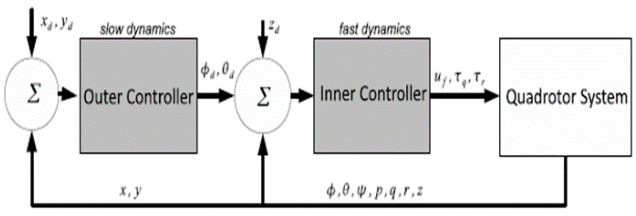


Fig. 13 Block diagram for FBL control strategy

FTC using FBL in case of failure on the actuator 2 is described in [15]. In the first simulation, only the inner controller is incorporated while the outer control is deactivated. The desired values for the roll and pitch are all set to 0 except for the altitude which is chosen to be$10m$. The simulation was run for $20s$ which is a sufficient time to reach hover.

As it can be seen from Fig. 14 the attitude angles $\emptyset$ and θ go to zero in less than $15s$ and with a smooth profile while the altitude z is quickly regulated to the desired value of $10m$.

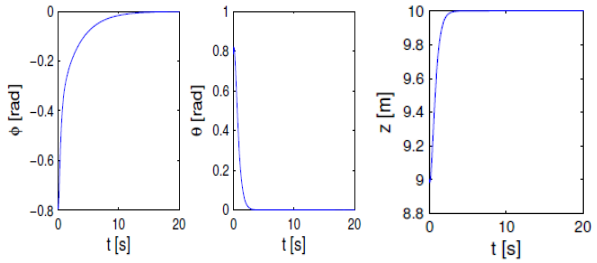


Fig. 14 System response with only inner controller applied

In the second simulation, both the inner controller and the outer controller are incorporated. The desired values for the roll and pitch are all set to 0 except for the altitude which is chosen to be10m. If the quadrotor does not reach the desired lateral and longitudinal position, then it is set to a ramp with negative slope until it becomes zero (landing procedure).

Figure 15 shows that the attitude angles $\emptyset$ and θ are stabilized, but this time there are oscillations due to the presence of the outer controller.

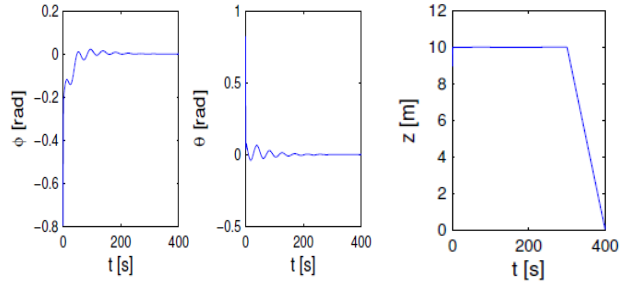


Fig. 15 Orientation angle stabilization with both inner and outer controller applied.

F. Sliding Mode Controller (SMC)

In SMC, the common task is to design a state feedback control law that maps the current state $X(t)$to the input u to stabilize the system around the origin $X = [0\ 0,..0]^T$. Whenever the system is started away from origin, it will return to it. SMC forces the system trajectories into a constrained subspace then holds them there so that they slide along with it.

In order to have a system track $\emptyset(t) \equiv \emptyset d(t)$, a sliding surface is defined as follows [16]:

$$s_\phi = \dot{z}_\phi + \lambda_\phi z_\phi \tag{7}$$

Where $z_\emptyset = \emptyset_d - \emptyset$ Then we have:

$$\dot{s}_\phi = \ddot{\phi} - \ddot{\phi}_d + \lambda_\phi \dot{z}_\phi \tag{8}$$

$$= \ddot{\theta}\ddot{\varphi}\frac{I_Y - I_Z}{I_X} + f_1 + \frac{l}{I_X}U_1 - K_4\dot{\phi} + g_1 - \ddot{\phi}_d + \lambda_\phi \dot{z}_\phi$$

The best approximation of U_1 for a continuous control law is obtained by setting $\dot{s}_\phi = 0$. The same steps for achieving the best U_1 approximation are used to extract U_2, U_3 and U_4

Main drawback of sliding mode control is the chattering due to the switching from the normal operation mode and the faulty mode.

Farid and co-researchers [16] describes the fault tolerance property of SMC and uses it in an FTC. The objective of their work is to land the quadrotor horizontally ($\emptyset = 0\ and\ \theta = 0$) when an actuator fault occurs. During fault free operation, signal U_2 that controls the pitch angle is equal to the difference of thrust of the motor1 and 2. When there is a fault in the motor 1, it is possible that it cannot provide sufficient thrust for control signal U_2 . To resolve this problem, the effect of motor1 in control signal U_2 is omitted and the control signal is only provided with the motor2.

A faulty condition is introduced after 7 seconds to rotor 1 and the performance of the quadcopter is observed without and with the SMC.

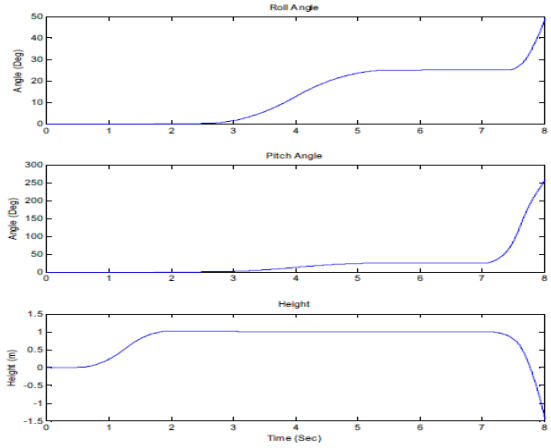


Fig. 17 Quadcopter performance in faulty situation with SMC

Li et al. [17] investigates both passive and active the fault-tolerance property of sliding mode controllers. Both simulation and experimental results using the Qball-X4 system model were performed. It was observed that the active FTC is more robust and shows better tracking performance in the presence of faults. A detailed analysis emphasizing the advantages and disadvantages of each type is discussed.

Studies presented in literature are believed to present the maximum fault value a passive and an active fault tolerant control SMC can handle. The SMC has not been tested for higher faults and severe actuator failures.

G. Model predictive control (MPC)

The MPC recalculates the control signal at each sampling time. Any change in the process dynamics is reflected into the control signal calculation [18]. The drawback of MPC is that it needs an almost explicit model of the system to calculate a stabilizing control signal. In addition, the abrupt changes in the model parameters, due to failure, requires a Fault Detection and Diagnosis (FDD) to provide information about the occurring faults to allow MPC to consider faults.

The system dynamics in presence of an actuator fault can be rewritten as:

$$x(k+1) = f(x(k), \alpha(k)u(k)) \quad (9)$$

Where $\propto$ is the fault parameter matrix capturing the fault severity. In derivations scalar $\propto_i$ denotes the amplitude of the fault in the i^{th} actuator. $\propto_i = 1$ denotes fault free actuator, $\propto_i = 0$ denotes a complete loss of actuator effectiveness and $0 < \propto_i < 1$ denotes a partial loss.

For a given time step k, MPC generates the input and state trajectories, by solving an optimization problem Q_k . The MPC controller applies only the first computed control input, $u_k(0)$ to the system and the time step is then incremented. The process is recursive as shown:

Given $x(0)$ and r^i , where x is the actual state vector and r^i is the reference state, then;

- $K = 0$
- Measurement (or estimate) $x(k)$
- Solve Q_k and generate u_k and x_k
- Apply u_k (0) to the system
- $K = K + 1$ and GOTO step2

An MPC strategy is proposed in [19], sacrificing the control on yaw. According to the simulations, the MPC method is able to get the quadrotor UAV in hover position. However, the controller cannot be implemented on hardware for experimental results since the angular velocities are very high and may cause problems therefore there is need to physically validate these simulation results. From Fig. 18 the yaw angle goes to infinity due to the aerodynamic drag of the vehicle, which was not taken into account in this work.

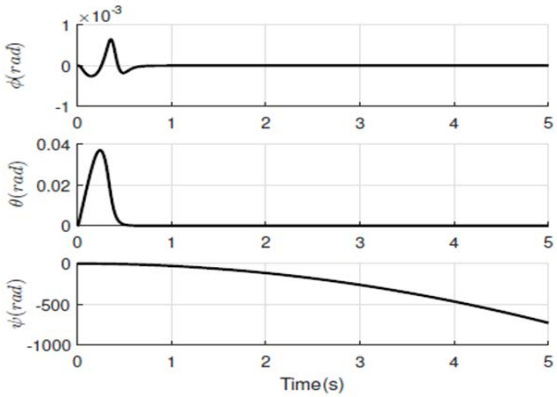


Fig. 18 Euler angles of the quadrotor UAV when rotor 1 is shut down.

The presented simulations showed that the roll and pitch angles were stabilized at the desired angles. However, MPC needs a relatively accurate model of the post-failure system to calculate a stabilizing control signal. The problem becomes more critical where the system dynamics is described by a nonlinear model [19]. This therefore calls for nonlinear parameter estimators such as Moving Horizon Estimation (MHE) and/or Unscented Kalman Filter (UKF) for reasonable online computation time as presented in [20].

H. Backstepping control

This is a Lyapunov stability based algorithm. The higher order nonlinear system is disintegrated systematically into a number of lower order subsystems. The lower order subsystems are arranged in cascaded form with each other to form overall closed loop system. Each lower order subsystem is stabilized by another lower order subsystem which generates a control input. This recursive loop keeps going until feedback control is achieved. This feedback control law is converted, through parameter estimation, into dynamic control law to accommodate the dynamic perturbations in parameters.

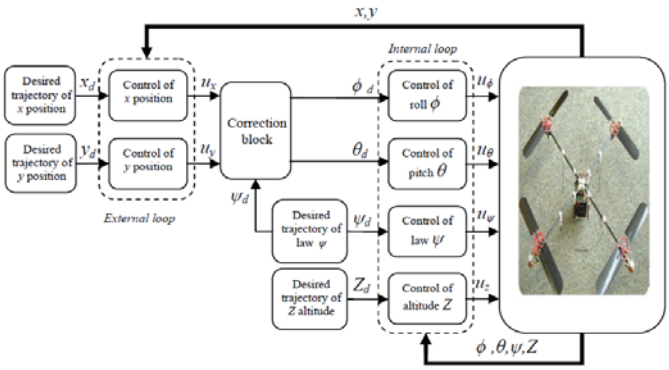


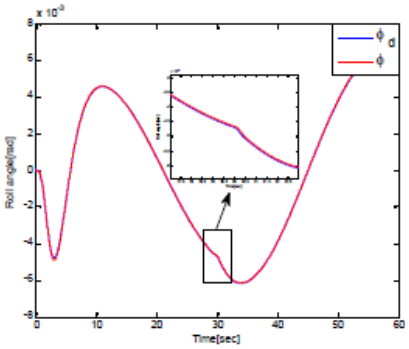
Fig. 19 Synoptic Scheme of the proposed control strategy [21]

The adopted control strategy in [21] describes a control strategy based on an internal loop and external loop. The internal loop controls the roll, pitch, yaw and altitude. The external loop includes control laws of positions x and$y$. The external control loop generates a desired roll (ϕ_d) and pitch (θ_d) through the correction block as follows:

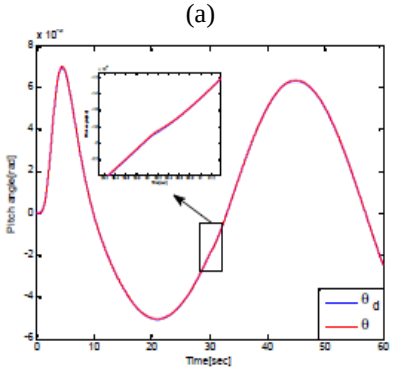
$$\phi_d = \arcsin(U_x Sin(\varphi_d) - U_y \cos(\varphi\_d)) \quad (10)$$

$$\theta_d = \arcsin(\frac{U_x \cos(\varphi_d) + U_y Sin(\varphi_d)}{\cos(\phi_d)}) \quad (11)$$

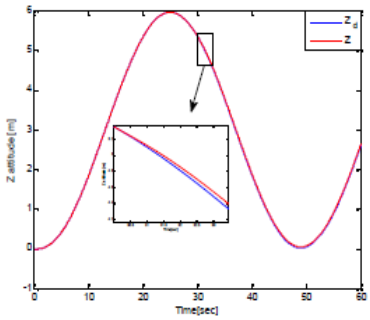
In Backstepping approach, a recursive algorithm is used to synthesize the control laws forcing the system to follow the desired trajectory in presence of actuator faults.



(a)



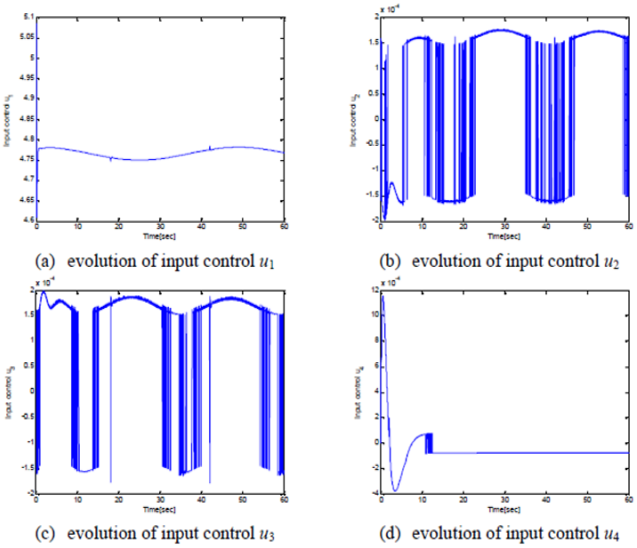
(b)



(c)

Fig. 20 (a), (b) and (c) shows tracking simulation results of trajectories along roll (ϕ), pitch (θ) and Z axis

The simulation results in [21] shows robustness towards stability and tracking even after the occurrence of actuator faults explaining the efficiency of the control strategy developed. However, Backstepping approach is affected by chattering effects on the inputs as shown in Fig. 21. There is therefore need to develop other control strategies in order to eliminate the chattering phenomenon in inputs control u_1, u_2, u_3, u_4, while maintaining the stability and the performances of this system, with implementation them on a real system.



(a) evolution of input control u_1 (b) evolution of input control u_2

(c) evolution of input control u_3 (d) evolution of input control u_4

Fig. 21 chattering effects on inputs

IV. CONCLUSION

It can be concluded that the discussed algorithms have advantages over each other. Therefore, it can be pointed out that the limitations discussed can be compensated by the advantages of the other algorithm. . It is therefore necessary to also look at Knowledge based and Hybrid Algorithms in future for comparison with most of the classical algorithms discussed in this review. This review acts as a stepping-stone into further work that will be conducted by the authors on design of Quadcopter Fault tolerant controller.

V. REFERENCES

[1] A. Girard, A. Howell, and J. Hedrick, “Border patrol and Surveillance missions using multiple Unmanned Air Vehicles,” in in 43rd IEEE Conference on Decision and Control (CDC), vol. 1, pp. 620 - 625, Dec 2004.

[2] S. Lee, S. H. Kang, and Y. Kim, “Trajectory tracking control of quadrotor UAV,” in in 11th International Conference on Control, Automation and Systems (ICCAS), pp. 281 – 285, Oct. 2011.

[3] S.Bouabdallah, R.Siegwart, “Full control of a Quadrotor” ,Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on San Diego, CA, USA .

[4] S. Bouabdallah, “Design and Control of Quadrotors With Application to Autonomous Flying”, PhD thesis, Ecole Polythechnique Federale De Lausanne, 2007.

[5] M. Numan ,“Controller Design for Attitude and Position Control of Quadrotor,” Master’s thesis, Capital University of Science and Technology, Pakistan, 2017.

[6] K. Cho, J.Shin and T.Kuc , “Design of Quadrotor Controller for Emergency Situation using Xplane” The 12th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI 2015) October 28~30, 2015 / KINTEX, Goyang city, Korea

[7] C. Balas, “Modelling and Linear Control of Quadcopter,” M.S. Thesis, Cranfield University, Cranfield, 2007

[8] Y.Zhang, “Concordia University Fault-Tolerant Trajectory Tracking Control of a Quadrotor Helicopter Using Gain-Scheduled PID and Model Reference Adaptive”,Annual Conference of the Prognostics and Health Management Society, 2011

[9] I.Sadeghzadeh, A.Mehta, Y.Zhang, “Fault-Tolerant Trajectory Tracking Control of a Quadrotor Helicopter Using Gain-Scheduled PID and Model Reference Adaptive Control”, Annual Conference of the Prognostics and Health Management Society, 2011

[10] E. Gopalakrishnan, “Quadcopter Flight Mechanics Model and Control Algorithms”, Master’s Thesis, Czech Technical University Prague, May 2017.
[11] R.Castillo, R.Lozano & A.Dzul, “Stabilization of a mini rotorcraft with four rotors. Control Systems”, IEEE, 25(6), p.45- 55. Available at: [Accessed March 3, 2018.
[12] T.Lan, Q. Fei , Q.Geng, “Control Algorithm Research of Quadrotor system based on 3-DOF Simulation Platform”, IEEE International Conference on Automation Quality and Testing Robots (AQTR). Beijing Institute of Technology. 2010
[13] S.Bouabdallah, S.Noth, R.Siegwart, “PID vs LQ Control Techniques Applied to an Indoor Micro Quadrotor”, Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004), Vol. 3, 28 September-2 October 2004, 2451-2456.
[14] A. Zulu, S.John, “A Review of Control Algorithms for Autonomous Quadrotors, Open Journal of Applied Sciences”, 2014, 4, 547-556.
[15] A.Freddi, “A Feedback Linearization Approach to Fault Tolerance in Quadrotor Vehicles”, Proceedings of the 18th World Congress The International Federation of Automatic Control Milano (Italy) August 28 - September 2, 2011.
[16] F.Sharifi, M.Mirzaei, W. Gordon, and Y.Zhang, “Fault Tolerant Control of a Quadrotor UAV using Sliding Mode Control”,Conference on Control and Fault Tolerant Systems Nice, France, October 6-8, 2010.
[17] Li, T., Y Zhang, B. Gordon, “Passive and active nonlinear fault-tolerant control of a quadrotor unmanned aerial vehicle based on the sliding mode control technique”, Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering 227(1): 12–23.
[18] J. M. Maciejowski, & C. N. Jones, “MPC fault-tolerant flight control case study: flight 1862”, IFAC Symposium on Safeprocess,Washington D.C., USA.
[19] L. Wang, T.Beumer, “Fault Tolerant Control of a Quadrotor Unmanned Aerial Vehicles”, Masters thesis, RMIT University, Tutor DBL (OGO) at Eindhoven,2016.
[20] A. Hojjat, H. Izadi, “ Fault Tolerant Model Predictive Control of Quad-Rotor Helicopters with Actuator Fault Estimation”, Preprints of the 18th IFAC World Congress Milano (Italy) August 28 - September 2, 2011.
[21] H. Khebbache, “Robust Stabilization of a Quadrotor Aerial Vehicle in Presence of Actuator Faults”, International Journal of Information Technology, Control and Automation (IJITCA) Vol.2, No.2, April 2012

A REVIEW ON ARTIFICIAL NEURAL NETWORK MODELS FOR SHORT TERM WIND POWER PREDICTION

Joseph N. Mathenge [1], D.K Murage [2], J.N Nderu [3] and C.M Muriithi [4]

***Abstract* —** There has been a growing global need to develop tools or models that are able to perform accurate wind power prediction in a power system. This arises from the fact that modern day power systems have inclined towards using renewable sources of energy; key among them being solar and wind. On one hand, solar power is relatively predictable and easily dispatchable while on the other hand, wind power is highly variable and intermittent hence making it have limited dispatchability. Current research is geared towards developing tools that can easily and accurately predict the short term wind power expected and hence make it a dispatchable resource. There have been many algorithms developed by researchers across the world to do short term prediction of wind power. However, one tool still remains supreme among all and that is the artificial neural network. This is due to its learning ability. This paper reviews the use of artificial neural networks in short term wind power prediction. The aim is to achieve accuracy that shall make wind power a fully dispatchable resource to ensure power systems gain from this free resource.

***Keywords*—** Artificial Intelligence, Artificial Neural Networks, Short Term Wind Power Prediction.

I. INTRODUCTION

THE adoption of the Kyoto Protocol in 1997 ignited a shift in the focus of countries from fossil fuels to renewable energy sources thus opening up one of the biggest research areas in science. The technology around renewables has and is still growing at a very high rate, something that has seen the developed as well as the developing nations embrace renewables more and more [1]. Wind has been the most widely accepted renewable energy source to be adopted by many power systems around the world [2]. Currently, the bulk of energy requirements in the world comes from fossil fuels and these are getting depleted at a very fast rate. On top of that they also have a huge negative impact on the environment in terms of their carbon footprint. These are among the reasons why the world is inclining towards investing in renewable sources of energy.

Wind energy is considered to be among the most economical sources of energy considering the fact that no fuel cost is incurred since wind occurs freely in nature. Due to its intermittent nature however, integrating it into the grid poses great challenges. This brings about the need for the development of highly accurate prediction methods to ensure reliability of the power system [3]. The prediction problem can either be a short term problem, medium term or a long term problem. The solution selected to solve this problem is also dependent on the time frame in question.

The most important among these for power system operators is the short term analysis since this allows power system operators make critical decisions about the running of the grid. The predicted output is compared with the actual output from a wind turbine or a wind farm and if the results are agreeable within a given margin of error then the method is considered accurate. In mature grids, power marketers are surcharged in the event that they underrate or overrate the quantity of wind power that shall be injected to the grid at any one time [1].

In this paper, the use of neural networks in short term wind prediction is critically analyzed to determine its effectiveness and how previous researchers have been able to enhance the tool.

II. REVIEW

A. Forecasting Time Horizons

There are four forecasting time scales according to S.M Lawan et.al, 2014 [4] as presented in Table I below:

Time Horizon	Range
Very short-term	Few seconds to 30 mins ahead
Short-term	30 mins to 6 hours ahead
Medium-term	6 hours to 1 day ahead
Long-term	1 day to 1 week or more ahead

Table I: Wind Prediction Time Scales

The most popular time horizon with researchers and power system operators is the very short term and the short term horizons. Both of these time horizons of interest are greatly involved in electricity market clearing as well as dispatch of wind power. Time is of the essence when it comes to working with short horizons and the method chosen to solve this problem should provide optimal solutions with minimal computation times. For the results to be acceptable, the predicted values of wind power need to be really close to the actual values to ensure efficient running of a power system with significant wind power penetration.

Joseph N. Mathenge, Department of Electrical Engineering, JKUAT (Phone: +254725096991; e-mail: jmathenge@ jkuat.ac.ke).

Prof D.K. Murage, Department of Electrical Engineering, JKUAT
Prof J.N Nderu, Department of Electrical Engineering, JKUAT
Prof C.M Muriithi, Department of Electrical Engineering, MUT

B. Forecasting Methods

There are four categories of wind speed/wind power forecasting:

- Persistence model – This was the pioneer model to be used in the wind speed/wind power prediction problem. It has been used as a benchmark from which all other developed models have to be compared to in order to establish their effectiveness. The persistence model assumes that the wind speed at time t = t+x is the same as wind at time = t. This argument only holds for very short term predictions. As the time frame of interest increases the accuracy of the persistence model reduces [4].
- Numerical Weather Prediction (NWP) model – This is mainly dependent on meteorological data for its accuracy. It is also referred to as the physical model of wind speed/power predictions. Its inputs are normally weather parameters that affect the wind speed at a given area. It involves complex mathematical models and hence have high computational times.
- Statistical model and ANNs – Are mainly used for short term prediction problems. They are easy to understand since they use the black box approach unlike the difficult mathematical models in NWP model. Mostly, historical wind power data alone is sufficient for use in this model. The more variables the data has, the better the relationship drawn between the inputs and predicted output and hence better performance of the model.
- Hybrid models – These involve the use of two or more models with the aim of solving a prediction problem. A statistical model can be used in conjunction with a physical model, combination of one model that solves the short term problem and another solving the medium term problem, combination of statistical models among many other hybridization forms that can exist. The level of hybridization is only limited by the creativity of the researcher.

ANNs have been widely used in prediction models for power systems and in other fields due to their learning ability – something that makes neural networks stand out when compared to any other tool. This paper focuses on reviewing the performance of ANNs in wind speed/wind power predictions in grids globally.

C. Artificial Neural Network (ANN):

ANNs are a non-linear mapping architecture that borrow from the neurons structure and operation of the human brain. The ANN is a robust tool of prediction, more so, for situations where data relationships are unknown and are seeking to be established. The forecasting of wind speed is such a problem. The relationship between one wind speed and the next is a non-linear function that cannot be accurately predicted. ANNs learn from any correlated patterns that are observed in the input data sets and set target values. As a result of this pattern, it is able to forecast the next expected outcome based on the input data sets and target values.

Training an ANN allows it to be used to predict the outcome of new independent input data. ANNs mostly deal with data that is considered vague and noisy which at times changes erratically. Hence, they are ideally suited for modeling of a wind power data prediction tool since such data is complex and often non-linear [5]. Neural networks perform pattern matching tasks that have a large number of highly interconnected nodes. This interconnection of neurons, just like in the human brain, gives ANNs the ability to learn and generalize training patterns from the training data. This strong learning capability of ANNs is the main advantage that makes it the best tool for use in wind speed prediction in a power system.

The structure of a neural network consists of three layers:

i. Input layer
ii. Hidden layer
iii. Output layer

The input layer accepts the input data to the ANN. This data is referred to as training data, or the simulation data. The hidden layer (or layers) is where modification and manipulation of the inputs takes place. It gives the neural network the ability to generalize the training data that is fed into it. Also, the interconnection of the neurons is defined in this stage. It is the hidden layer that determines the quality of the solution provided by the neural network. The final layer is the output layer and this gives a single output from the ANN.

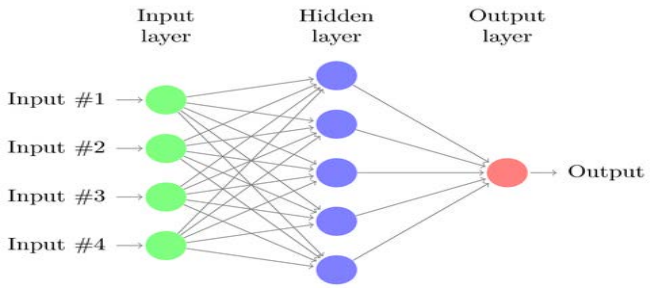


Fig. 1. Architecture of an ANN [5].

D. Previous Research

The history of ANNs dates back to 1959 when Rosenblatt introduced the single layer perceptron network. This would later be improved in 1962 by Minsky and Papert who introduced the multilayer perceptron network after proving the inadequacies of the single layer perceptron network. To date, the multilayer perceptron (MLP) network is the most common [6]. In 1968, Rumelhart et al introduced the back propagation network (BPN) for use in training of the MLP model. BPN provides a very efficient method of updating the weights of the neural network and this is key to ensure the accuracy of the ANN. A schematic of the BPN is as shown in Fig. 2 below:

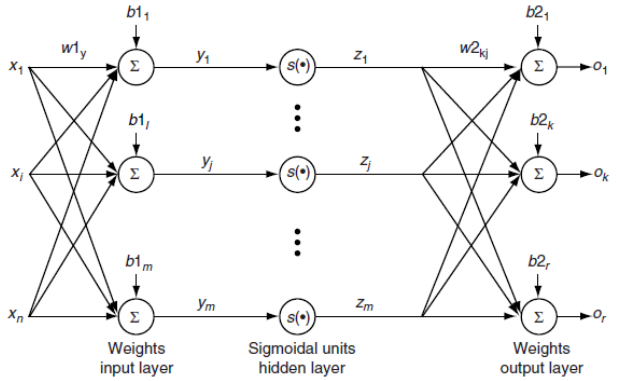


Fig. 2. Schematic of a BPN [6].

Other neural network models that have been developed are the radial basis function network by Powell in 1985 and the recurrent neural network by Elman in 1990. They have both been proved to have excellent performance in dealing with non-linear approximation problems [6].

The output of a wind turbine is given as [7]:

$$P_{wind} = \frac{1}{2}\sigma A v^3$$

Where:
σ – Air density in kg/m^3
A –Swept area of the wind turbine/Cross section area of the turbine
v – Upstream wind speed

Variables that fall under inputs to wind prediction tools fall within the range of parameters that affect generation from wind plants. They include: air density, wind speed, humidity, amongst other variables. Accuracy is analyzed based on the ME, MSE, MAE, RMSE, comparison with the persistence model etc. to establish how the proposed system performs based on the actual values of wind power and comparatively with other prediction tools. ANNs can either be used alone, in a hybrid state or in conjunction with other algorithms.

In 2006, C. Potter and M. Negnevitsky implemented an ANFIS model for use in wind power forecast in Tasmania with wind speed and wind direction as the inputs [8]. The results obtained gave an error of 4% and were limited to accuracy within a 2.5 min ahead forecast only. When the same data was analyzed using persistence, a 30% error was observed hence showing the robustness of a hybrid form of neural networks (ANFIS).

In 2007, a recurrent fuzzy neural network was developed by Barbounis and Theocharis [9] to perform multistep forecasts from 15mins to 3 hours ahead. The combination of fuzzy and neural networks was observed to improve the results obtained. ANFIS combines the advantages of fuzzy and neural networks hence giving more accurate results with reduced computation times. Palomares *et al* in [10] compares the performance of the ARIMA and neural network models. The back propagation model with Levenberg Marquardt training comes out as more powerful compared to the ARIMA model. It, however, consumes more computation capacity in form of memory.

The neural network was coupled with the Markov chain model in [11], by Pourmousavi in 2011, for very short term wind prediction achieving a MAPE of 3.1439. The persistence model posted a MAPE of 3.6821 for the same data set. Again, it is observed that modifying neural networks with the Markov chain further improves its performance in short term wind power prediction.

In 2016, Vijendra Singh [7] used a feed forward neural network with supervised learning using back propagation is used. The ANN structure used is shown in Fig. 3 below:

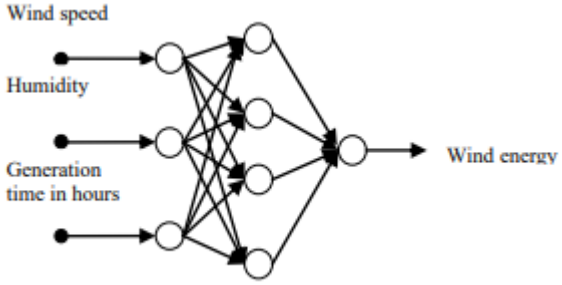


Fig. 3. Feed Forward Neural Network Proposed by [7]

Training of the network was done in batch mode in ranges from 10 mins to 60 mins. The MSE was set at 0.0001. After 300 iterations, the MSE stabilized at 0.0070 hence proving that the ANN used alone is a powerful prediction tool. It was observed that as the forecast period increases from 10 mins to 60 mins, the error grew hence confirming that ANNs are better suited for short term predictions.

In 2016, Hasan Masrur in [12] used hourly wind mean data for prediction. Using the nntool in MATLAB with the Levenberg-Marquardt feed forward back propagation. Various scenarios were tested and gave a MAPE of between 1.1 and 13.7. According to the MAPE criterion of determining precision, a MAPE value of less than 10% indicates high prediction, 10%-20% good prediction and over 50% inaccurate prediction. From the MAPE values obtained by [12], the neural network can again be confirmed to be a good tool for short term prediction.

ANFIS in wind power forecast was compared with Genetic Algorithm Back Propagation Neural Network and ANNs in [13] and results in Table II were noted:

Criterion	ANN	GA_BPNN	ANFIS
MAE	31.87	29.14	28.39
MAPE	5.86	4.59	4.45
RMSE (kW)	48.45	46.83	46.06
MSE (kW)	2346.87	2192.35	2121.5

Table II: Results of a comparative analysis of ANFIS, ANNs and GA-BPNN in short term wind power prediction.

The robustness of ANNs both in its original state and in hybrid form is noted. The values of errors obtained in the three methods being compared are in the same neighborhood of each other. However, it is noted that hybrid forms of ANNs perform comparatively better.

In [14], research is done on the accuracy of neural networks in wind power prediction by testing three different scenarios as shown in Fig.4 below:

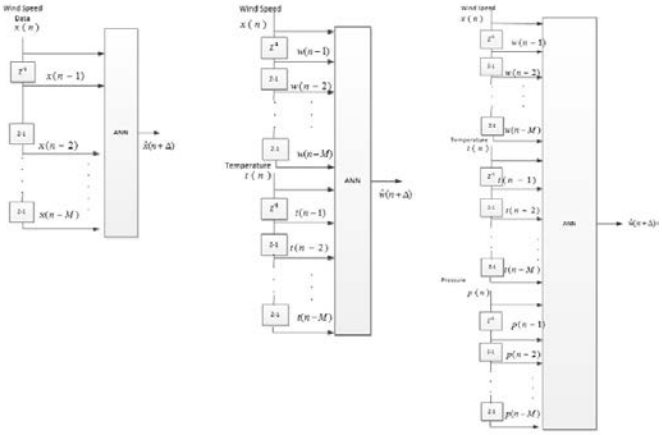


Fig. 4: Scenario 1 – Wind speed data only, Scenario 2 – Wind speed and temperature data, Scenario 3 – Wind speed, temperature and pressure data [14].

For each scenario, the RMSE and MAE values were computed for forecast lead time, Δ of 30 and 90. The results obtained are as tabulated below:

Forecast lead time, Δ = 30		
Scenario	RMSE	MAE
1	0.6508	0.5046
2	0.6502	0.5040
3	0.6494	0.5032

Table III – RMSE and MAE performance for three scenarios in [14] with forecast lead time Δ = 30.

Forecast lead time, Δ = 90		
Scenario	RMSE	MAE
1	0.6940	0.5537
2	0.6903	0.5470
3	0.6759	0.5360

Table IV – RMSE and MAE performance for three scenarios in [14] with forecast lead time Δ = 90.

It can be noted in Tables III and IV that the performance of scenarios 2 and 3 outperform the results obtained for scenario 1. This serves to prove that in as much as the neural network is a powerful tool, there is also the need process the data being fed into the network more. The more shallow or limited the data fed into the neural network is, the less accurate the solution obtained as is observed in scenario 1. When more variables that affect wind speed are introduced, the solution becomes more optimal since the neural network is able to draw more relationships from the more information fed into it.

The results reviewed above are just but samples in a huge matrix of research that has been done to determine the robustness of neural networks in wind speed and wind power predictions across the globe. In a review by [6], neural networks are seen to have been used for the last two decades in solving the wind forecast problem with great success. ANNs post more accurate results, competitive computation times, and minimal errors compared to non – ANN models. When hybridized or used in combination with other superior models, the results are improved with better accuracy and reduced computation times [6].

III. DISCUSSION ON PERFORMANCE OF ANN MODELS

Neural network models have proved to be very robust tools in handling non-linear problems. Wind forecasting is a non-linear problem and neural networks have posted very significant success in solving it. Optimal solutions have been obtained using ANNs alone and even better solutions obtained when the ANN is hybridized or used in conjunction with other superior algorithms. Neural network models are seen to have very competitive computation times, accurate results with minimal margins of error and proves that wind speed and wind power can be predicted in advance for use in optimizing power system operation and planning.

The approach that the neural network is employed in to solve the wind forecast problem is only limited to the imagination of the researcher. A lot of work can still be done using this robust tool by improving its performance through hybridization or combination with other tools to improve the results further. In addition to this, pre-processing of data before feeding it into the neural network model is a great determinant of the performance of the model. As is the case with any computer system, the garbage in-garbage out principle still applies here. If the input to the model is flawed, the neural network shall draw wrong connections in the data leading to an output that is not optimal.

IV. CONCLUSION

Bulk wind power injection into the grid has brought to the spotlight the importance of prediction tools in power systems. No other source of energy has ever given power system operators a more difficult time to integrate into the grid than renewables with stochastic nature. The wind harnessing technology has matured since its inception and prediction tools are becoming better by the day with minimal computation times and accurate results with smaller margins of error. Wind is bound to be converted from a non-dispatchable resource to a dispatchable resource with precise forecasts and planning making grids across the word benefit from this free resource.

V. REFERENCES

[1] Abinet T., J.H Zhang, D.H Zheng and Dereje S., "Short-Term Wind Power Forecasting Using Artificial Neural Networks for Resource Scheduling in Microgrids," *International Journal of Science and Engineering Applications*, vol. 5, no. 3, pp. 144 - 151, 2016.

[2] IEC, "Grid integration of large-capacity Renewable Energy sources and use of large-capacity Electrical Energy Storage," International Electrotechnical Commission, Geneva, 2012.

[3] IEC, "Grid Integration of Large Capacity Renewable Energy Sources and use of Large Capacity Electrical Energy Storage," in *International Electrotechnical Commission*, Geneva, 2012.

[4] S. M. Lawan, W. A. W. Z. Abidin, W. Y. Chai, A. Baharun and T. Masri, "Different Models of Wind Speed Prediction; A Comprehensive Review," *International Journal of Scientific & Engineering Research*, vol. 5, no. 1, pp. 1760 - 1768, 2014.

[5] E. Pratheepraj, Anuj Abraham, S.N. Deepa, and V. Yuvaraj, "Very Short Term Wind Power Forecasting Using PSO-Neural Network Hybrid System," *Springer-Verlag Berlin Heidelberg*, vol. III, pp. 503 - 511, 2011.

[6] K. Gnana Sheelaa and S.N. Deepab, "A Review on Neural Network Models for Wind Speed Prediction," *Wind Engineering*, vol. 37, no. 2, pp. 111 - 124, 2013.

[7] Vijendra Singh, "Application of Artificial Neural Networks for Predicting Generated Wind Power," *(IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 7, no. 3, pp. 250 - 253, 2016.

[8] C. Potter and M. Negnevitsky, "Very short-term wind forecasting for Tasmanian power generation," *Proc. IEEE Trans. Power Syst*, vol. 21, no. 2, pp. 965 - 972, 2006.

[9] Barbounis and Theocharis, "Locally Recurrent Neural Network for Wind Speed," in *Elsevier Information Sciences*, 2007.

[10] J.C Palomares, J.J. G. de la Rosa, J.G. Ramiro, J. Melgar, A. Aguera and A. Moreno, "ARIMA Vs Neural Network for Wind Speed Prediction," in *Int. conf. on intelligence for Measurement Systems and Applications*, 2009.

[11] S. A. Pourmousavi Kani, M. M. Ardehali, "Very Short-term Wind Speed Prediction: A New Artificial Neural Network-Markov Chain Model," *Energy Conversion and Management*, vol. 52, no. 1, pp. 738 - 745, 2011.

[12] Hasan Masrur, Meas Nimol, Mohammad Faisal and Golam Mostafa, "Short Term Wind Speed Forecasting Using Artificial Neural Network: A Case Study," *IEEE*, 2016.

[13] Yordanos Kassa, J. H. Zhang, D. H. Zheng and Dan Wei, "Short Term Wind Power Prediction Using ANFIS," *IEEE International Conference on Power and Renewable Energy*, pp. 398 - 393, 2016.

[14] Ümmühan Basaran Filik and Tansu Filik, "Wind Speed Prediction Using Artificial Neural Networks Based on Multiple Local Measurements in Eskisehir," in *Elsevier*, 2016.

A Review of Load Flow Methods in Analysis of Power Distribution Systems

Simon N. Chege, David K. Murage and Peter K. Kihato

***Abstract*—** Power systems consist of four important segments namely, generation, transmission, distribution and utilization. Each of these segments require to be planned and operated securely in order to maintain a given frequency and voltage level. In order to carry out effective planning, operation, optimization and control, load flows are important for determining the state variables of these networks. With proper method of load flow, the planner would be able to determine network problems such as voltage stability, network power loss and transient stability therefore operating the network securely and economically. In this paper, a review of methods of load flows used in analysis of distribution network has been carried out. The review seeks to highlight strengths and weaknesses of different load flow methods while studying distribution networks.

***Keywords*—** Distribution Network, Load Flow methods.

I. Introduction

POWER distribution networks form the link between transmission network and the consumer. Distribution networks have unique characteristics that differentiate them from transmission network. These characteristics include unbalanced distributed loads, multiphase unbalanced operation, large number of nodes, high resistance-to-reactance (R/X ratios) of the feeders and distributed generation [1].

A. Distribution Network components

Distribution networks consists of distributed feeders, distributor and service mains. Distributed feeders are conductors that connect a substation to the point where power is to be distributed. There are no tapping made on the distributed feeder and therefore the current in a feeder remains constant. A distributor is a conductor where tapping are made to the consumer. Current in a distributor varies due to tapping. A service main forms the link between the consumer terminal and the distributor.

A. Types of Distribution System Configurations

Depending on the feeder configuration, distribution networks can be classified into four categories namely radial distribution system, parallel feeders, ring main and meshed systems.

1) Radial Distribution systems

Radial Distribution Networks are the most common form of distribution system. This is because they are easy to construct, have a relatively simple protection scheme and are not expensive. The configuration of radial distribution system is as shown in fig. 1.

The disadvantage of these systems is that when a fault occurs on a feeder, the consumers connected on that feeder are disconnected during the entire period of the fault. They therefore are less reliable.

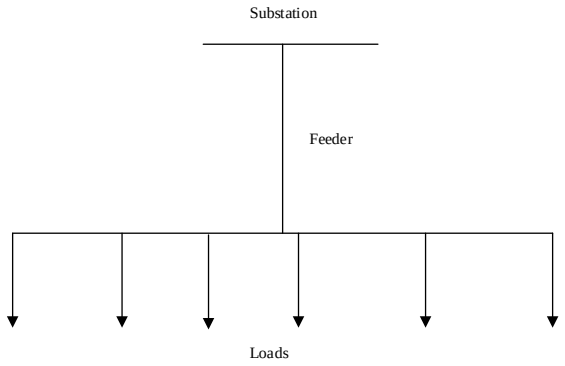


Fig. 1: Line diagram of a radial distribution system

2) Parallel Feeders

In this system, two parallel feeders run to the distributor through different routes. The aim is to increase reliability of the network in that when there is a fault in one feeder, the supply is maintained using the other feeder. This network however is more expensive than the radial distribution network. Fig. 2 shows a line diagram of a parallel feeder system.

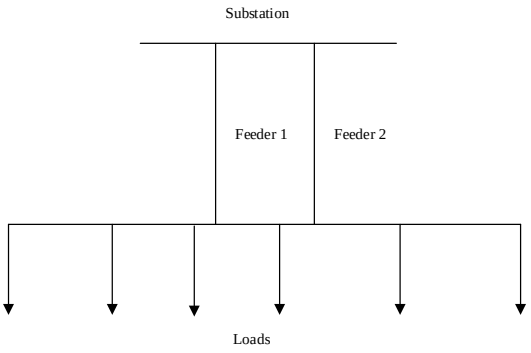


Fig. 2: Line diagram of a parallel feeder system

S. N. Chege, Department of Electrical and Electronic Engineering, JKUAT (phone: +2540724878860; e-mail: snchege@jkuat.ac.ke).

D.K. Murage, Department of Electrical and Electronic Engineering, JKUAT (e-mail: dkmurage25@yahoo.com).

P. K. Kihato, Department of Electrical and Electronic Engineering, JKUAT (e-mail: pkihato@eng. jkuat.ac.ke).

3) Ring main Feeders

In this configuration, the start and end of a feeder are at the same location. This increases liability because if there is fault on one end of the feeder, the supply is maintained through the other end. The configuration of this network is as shown in fig.3.

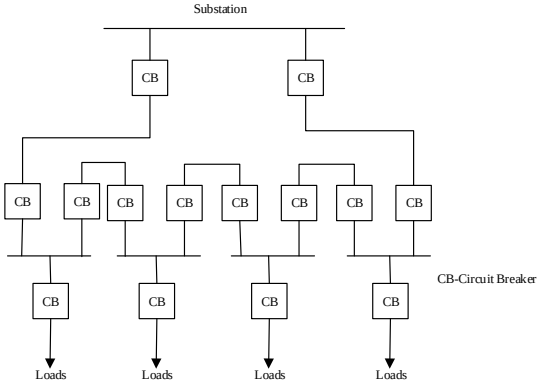


Fig. 3: Ring main feeder configuration

4) Meshed distribution systems

In meshed networks multiple paths are available between multiple points in the network. Power flow is split along several paths between any two points in the network. If a fault occurs at a point in the network, power flow reroutes to another path.

This type of network is the most reliable but also most complex and therefore difficult to analyze. This network is also the most expensive owing to its complexity and use of more conductors. Fig. 4 shows the configuration of a meshed network.

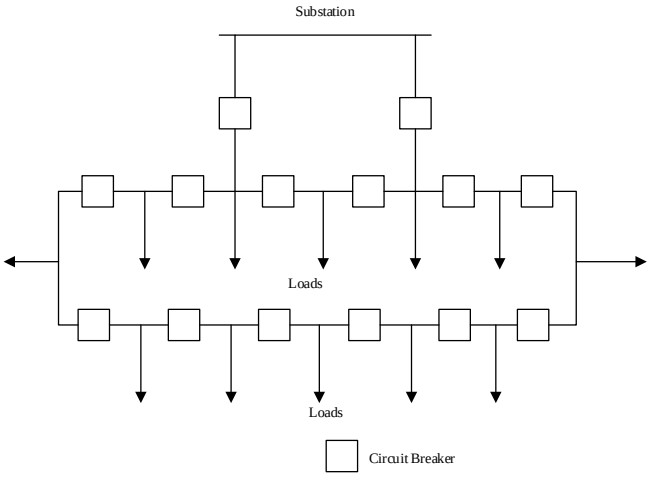


Fig. 4: Meshed distribution network

II. METHODS OF LOAD FLOW IN DISTRIBUTION NETWORKS

The choice of method of power flow is determined by the value of ratio of R/X of the network, presence of distributed generation, multiphase power flow and unbalanced loads [2].

Distribution networks have high R/X ratio and operate in unbalanced conditions. In addition, distribution networks are increasingly becoming active due to presence of distributed generation and as such new techniques of load flows are being adopted.

This section gives a review of various techniques that researchers have applied to solve load flows in distribution networks.

A. Forward and Backward Sweep Method

This method has been used in [3] to solve load flow for IEEE-33 bus radial distribution system. The method involved two steps, which are done iteratively, namely forward sweep and backward sweep.

In the forward sweep, voltage drop calculation is done. In addition, power flow and nodal voltages are updated in a forward direction starting from branches in the first layer towards those in the last layer. The effective power in each branch is held constant to the value obtained in the backward sweep during this step.

The backward sweep starts with branches in the last layer moving towards the branches connected to the root node. Effective power flows in each branch are obtained by considering the node voltages of previous iteration. The voltages obtained during the forward sweep are held constant during this step and updated power flows in each branch are transmitted backwards along the feeder using backward path.

The forward and backward sweep method has three variants determined by the quantity calculated during the backward sweep of each iteration. These include:

i) The current summation method where the branch currents are evaluated according to the equation (1).

$$J_L(k) = -I_{L2}(k) + \sum_{from\,node\,L2}^{currents\,in\,branches\,emanating} \quad (1)$$

where, $J_L(k)$ is the current in branch L at iteration k,

$I_{L2}(k)$ is current injection at node L and

$L = b, b-1, b-2, \Lambda\ ,1$

This is the direct application of the Kirchhoff's Current Law.

i) The power summation method where the power flows in the branches are evaluated.

ii) The admittance summation method where, node by node, the driving point admittances are evaluated.

The advantage of this method is that the Jacobian Matrix is not required. In addition, this method is suitable for radial and weakly meshed networks with a high R/X ratio. It also has a fast convergence [4].

B. Direct method(BIBC/BCBV matrix Method)

This method has been applied in [5] to solve load flow for a radial distribution system. It uses three steps namely, equivalent current injection, formulation of BIBC matrix and formulation of BCBV matrix.

1) Equivalent current injection

During this step, the current injection at bus i during the k^{th} iteration is computed as shown in equation (2).

$$I_i^k = \left(\frac{P_i + jQ_i}{V_i^k} \right). \quad (2)$$

where,

I_i^k -is equivalent current injection at the k[th] iteration for i[th] bus

P_i - is the real power at i[th] bus

Q_i - is the reactive power at i[th] bus

V_i^k - is the bus voltage at the k[th] iteration for i[th] bus

2) Formulation of BIBC matrix

This step involves forming current injection equations using the Kircchoff's Current Law for the radial distribution network and then writing the branch current as a function of the equivalent current injections. The relationship is given by equation (3).

$$[B] = [BIBC][I] \quad (3)$$

where,

$[B]$-Branch Current Matrix

$[BIBC]$-Bus injection-Branch Current Matrix

$[I]$-Bus Injection Matrix

3) Formulation of BCBV matrix

The BCBV matrix is represents the relationship between the branch currents and the bus voltages. Using the BCBV matrix, the respective variation of the bus voltages which is generated by the variation of the branch currents is established directly. This relationship is represented by equation (4).

$$[\Delta V] = [BCBV][B] \quad (4)$$

where,

$[\Delta V]$-Variation of bus voltage matrix

$[BCBV]$-Branch Current-Branch Voltage matrix which consist of line impedance parameters

$[B]$-Branch Current matrix

This method eliminates formation of Jacobian matrix and therefore its implementation is less time consuming. It is effective for analysis of radial networks.

C. Implicit Z_{BUS} Gauss Method

This method has been used in [6] to solve load flow flow for a three phase unbalanced radial distribution network. This method works on the principle of superposition. The voltage at each bus arises from the contribution of source bus voltage and the equivalent current injections.

When using the superposition principle, only one type of source is considered at a time while determining the bus voltages. That is to mean that when the slack bus is connected, all current injections are disconnected and vice versa.

The current injection at bus i at k[th] iteration is given by equation (2). The branch current vector and bus voltage vector at iteration k are given by equation (5) and equation (6) respectively.

$$I^{(k)} = Y\Delta V^{(k)} = LU\Delta V^{(k)} \quad (5)$$

$$V^{(k+1)} = V_{NL} + \Delta V^k \quad (6)$$

where,

Y - is the admittance matrix

$\Delta V^{(k)}$ -is the vector of voltage deviation of the k[th] iteration

LU -is the triangular factorization of the admittance matrix

V_{NL} -is the no load node vector matrix taken as equal to source node

The Z_{BUS} method is solved by equations (2), (5) and (6) iteratively.

Patil and Kurkani [6] concluded that the Z_{BUS} method performed better in execution time and rate of convergence than the conventional Newton Raphson method because it does not require computation of the elements of Jacobian. From this paper, the authors demonstrated the effectiveness of this method in solving load flow for unbalanced radial distribution systems.

D. Loop impedance method

The concept of loop impedance matrix has been used in [7] to solve power flow of three phase unbalanced radial distribution system. This method is based on graph theory where basic loop incidence matrix **C** and branch-path incidence matrix **K** of a connected graph are used to describe the system.

1) Basic Loop Incidence matrix

A distribution network is described by n nodes, e elements, b branches and l links. The number of branches is given by:

$$b = n - 1 \quad (7)$$

The number of elements is given by:

$$e = b + l \quad (8)$$

A basic loop incidence matrix **C** of a directed graph is an e x l matrix with elements of different dimensions according to number of phases of elements. The number of links and number of basic loops is the same.

Matrix **C(*i,j*)= +U** if element *i* is incident to and directed in the same direction as the j^{th} basic loop. Matrix **C(*i,j*)= -U** if element *i* is incident to and directed in the opposite direction as the j^{th} basic loop. **U** is a unit matrix whose dimensions correspond to the number of phases of element ***i***.

2) Branch path incidence matrix

The branch path incidence matrix represents the incidence of branches to paths in a tree, where a path is directed from a bus to the reference bus. The elements of branch path incidence matrix **K**(*i,j*)= **+U** if the branch *i* is in the path from bus *j* to the reference node and directed in the same direction. **K**(*i,j*)= **-U** if the branch *i* is in the path from bus *j* to the reference node and directed in the opposite direction.

Once the incidence matrices are formed, they are combined with the primitive network matrices to completely describe the system. In [7], loop impedance matrix $\mathbf{Z}_{loop}$ is obtained by combining the primitive impedance matrix |**Z**| and loop incidence matrix as shown in equation (9).

$$\mathbf{Z_{Loop}} = \mathbf{C}^t|\mathbf{Z}|\mathbf{C} \quad (9)$$

The network equation are obtained by equation (10).

$$\bar{V}_{Loop} = \mathbf{Z_{Loop}}\bar{I}_{Loop} \quad (10)$$

where $\bar{V}_{Loop}$ is the basic loop voltage vector, $\bar{I}_{Loop}$ is the vector of basic loop currents and $\mathbf{Z_{Loop}}$ is the loop impedance matrix.

This method has a fast convergence for large unbalanced radial distribution networks. In addition, it does not require formation of admittance matrix and therefore offers less computation time.

E. Newton Based Methods

Newton based methods are modified forms of the conventional Newton Raphson power flow solution. The modification is necessary because of the unique characteristics of distribution networks which affect the convergence of conventional Newton Raphson power flow method.

In [8], Polar Current Mismatch Version has been used to solve a three phase power flow problem in distribution network. This version is obtained using the current mismatch functions in polar co-ordinates.

Other versions include Cartesian current mismatch, complex current mismatch, Cartesian power mismatch and complex power mismatch which are detailed in [9]. Newton Based Methods have the disadvantage of using Jacobian Matrix which increases the computation time especially for large distribution networks.

III. CONCLUSION

This paper has reviewed methods used for load flow solution in distribution systems. These methods include forward and backward sweep method, implicit Z_{bus} Gauss Method, BIBC/BCBV matrix method, Newton Based Methods and Loop impedance method.

Newton Based methods use Jacobian matrix which requires modification in order to adapt it to the unique characteristics of a distribution system which include high R/X ratio and unbalanced loads. The other methods exploit the topological structure of the distribution system and therefore reduce the number of equations. This therefore means that the computational burden is reduced.

It can therefore be concluded that in solving load flow in distribution system, methods that exploit topological structure of the distribution network are more suited than the conventional load flow solution methods such as Newton Raphson.

REFERENCES

[1] K. Maya and E. Jasmin, "A Three Phase Power Flow Algorithm for Distribution Network Incorporating the Impact of Distributed Generation Models," Smard Grid Technologies, vol. 21, p. 326 – 331, 2015.

[2] B. Sereeter, K. Vuik and C. Witteveen, "Newton Power Flow Methods for Unbalanced Three-Phase Distribution Networks," Energies, vol. 10, no. 10, pp. 1658-1678, 2017.

[3] R. Michline and S. Ganesh, "Power Flow Analysis for Radial Distribution," International Journal of Electrical and Computer Engineering System Using Backward/Forward Sweep Method, vol. 8, no. 10, pp. 1621-1625, 2014.

[4] M. Nanghoguina, C. Muriithi and W. Wekesa, "Load Flow Analysis for Radial Distribution Networks Using Forward BackwardSweep Method," Journal of Sustainable Research in Engineering , vol. 3, no. 3, pp. 82-87, 2016.

[5] L. Vijay, K. Manish and R. Bajpai, "A Comparative Analysis of Distribution System Load Flow for 33-Bus System," International Journal for Electrical and Electronics Engineers, vol. 8, no. 1, pp. 1011-1021, 2016.

[6] G. Patil and S. Kurkani, "Performance Assessment of Load Flow Techniques for Unbalanced Distribution Systems," in National Power Systems Conference, Madras, 2004.

[7] H. Chen and C. Yang, "Loop frame of reference based three-phase power flow for unbalanced radial distribution systems," Electric Power Systems Research, vol. 80, pp. 799-806, 210.

[8] S. Baljinnyam, V. Kees and C. and Witteveen, "Newton Power Flow Methods for Unbalanced Three-Phase Distribution Networks," Energies, vol. 10, pp. 1658-1678, 2017.

[9] B. Sereeter, C. Vuik and C. Witteveen, "On a comparison of Newton-Raphson solvers for power flow problems," Delft University of Technology, Delft, 2017.

Distribution Generation and Capacitor Placement in Distribution Systems

Simon N. Chege, David K. Murage and Peter K. Kihato

***Abstract*—** Distribution systems form a critical part of the power system by linking the consumer to the transmission system. Distribution systems are extensive, complex and require adequate planning to ensure reliability and reduce installation and operation cost generated by both voltage instability and network power loss. High power losses and voltage instability are the main challenges facing power distribution companies. These challenges are mitigated by capacitor and distributed generation placement in the distribution network. The effectiveness of these measures is greatly dependent on optimally placing and sizing these components within the distribution network and also, how their operation is coordinated. Due to the complexity of the distribution networks, planning of placement becomes a complex task. It therefore requires network planners to develop methods that optimally place capacitors and distributed generations (DGs) in distribution networks. In this paper a novel way of DG and capacitor placement is proposed. The method uses Voltage Stability Index to find the optimal location of DGs and Capacitors. Evolutionary programming algorithm is employed to determine the sizes of the DGs and Capacitors to be placed at the identified locations. The aim is to enhance the voltage stability of the radial distribution network. This method is tested on the IEEE-33 bus radial distribution network. Simulation is carried out in MATLAB.

***Keywords*—** Capacitor Placement, Distributed Generation, Evolutionary Programming, Voltage Stability

I. Introduction

DEMAND for electric power has greatly increased due to economic growth and increasing population especially in developing countries. As a result of this, distribution systems are being operated close to their maximum limits of voltage stability and power carrying capacity. In addition, distribution systems have changed from passive networks to active networks due to increased proliferation of distributed generation [1]. Increased proliferation of distributed generation has resulted in a number of adverse effects. These effects include voltage variations, degraded protection, altered transient stability, bi-directional power flows and increased fault level. Voltage variation has been addressed as the most dominant impact of distributed generation [2].

Voltage stability is a requirement for the secure operation of distribution systems. Proper planning of distributed generation and their control strategies determine the voltage stability situation of distribution system [3] [4]. The planning aspect involves proper location and sizing of the distributed generation (DG) together with other reactive power sources in the distribution network. Control aspect involves the coordinated operation of these DGs together with conventional voltage and reactive power devices [5].

With the increased use of Distributed Generation, it has become critical to incorporate them in distribution system planning. The distributed generation problem has been a key area of research in the recent past. Different researchers have addressed distributed generation planning in different ways. The objective functions used by researchers in distributed generation planning include power loss minimization, reliability enhancement, minimization of operational and investment cost and voltage stability enhancement [6].

Researchers have used different methods in DG placement problems. These methods include analytical methods, numerical methods and heuristic methods. Heuristic methods have been found to work well for large and complex optimization problems such as DG and capacitor placement problem [7]. Heuristic methods that have been used in DG placement planning include particle swarm optimization [8] [9], bacteria foraging algorithm [10], differential evolution algorithm [11] and ant colony algorithm [12].

In this paper evolution programming algorithm is used to optimally place DGs and capacitors in order to enhance voltage stability.

II. PREVIOUS WORK

A. Voltage Stability

Voltage stability is the ability of power system to sustain acceptable voltage levels under normal operating conditions and after occurrence a disturbance [13]. Voltage stability is usually represented by P-V curves, Q-V curves and stability indices. Fig. 1 represents P-V curves for different load power factors.

The point of voltage collapse (PoVC) is the nose of the P-V curve. This is the point at which an increase of load leads to rapid voltage drop of the power system and consequently voltage collapse or network collapse. Voltage collapse usually occurs in heavily loaded systems that do not have sufficient local reactive power sources and consequently cannot provide secure voltage profile for the system [14]. The Q-V curves shown in fig. 2 show variation of receiving end voltage with

S. N. Chege, Department of Electrical and Electronic Engineering, JKUAT (phone: +2540724878860; e-mail: snchege@jkuat.ac.ke).

D.K. Murage, Department of Electrical and Electronic Engineering, JKUAT (e-mail: dkmurage25@yahoo.com).

P. K. Kihato, Department of Electrical and Electronic Engineering, JKUAT (e-mail: pkihato@eng. jkuat.ac.ke).

variation in load reactive power for different real power loads. The region of the locus of knee point represents the stable region.

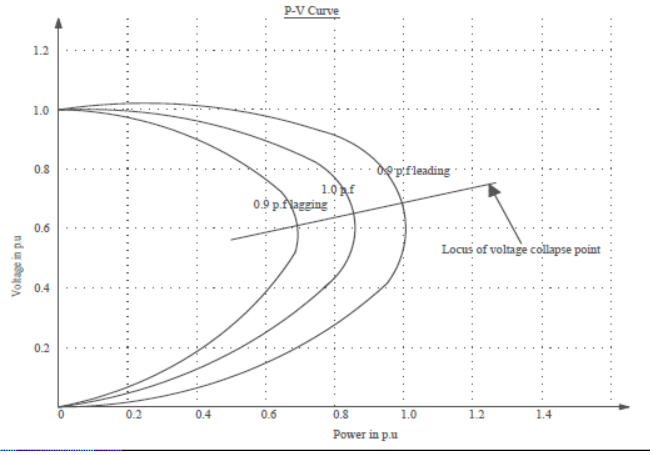


Fig. 1 P-V Curves for different load power factors

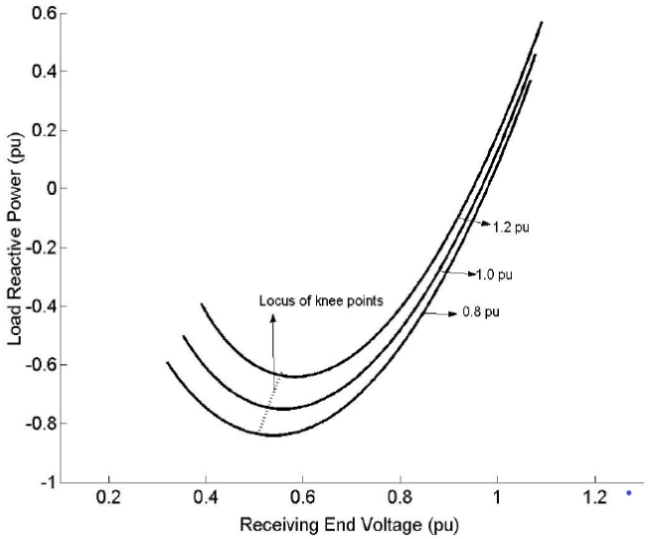


Fig. 2 Q-V Curves for different load real powers

When using the Q-V method, the sensitivity of voltage to changes in reactive power at a given bus is given by the slope of the Q-V curve. If the V-Q slope of the i^{th} bus is positive, the system is voltage stable and if negative the system is voltage unstable. Other methods used for steady state voltage stability analysis are modal analysis and sensitivity analysis. These method use the Power Flow Jacobian that is obtained by solving a set of equations linearized about a given operating point.

The Jacobian is evaluated for singularity to determine the maximum loadability of the power system. The main disadvantages of these techniques is that they require considerable computation efforts and are time consuming especially for a large network.

Recently, researchers have developed indices for voltage stability analysis in power systems and more so for analysis in distribution systems. Distribution networks are large and complex and therefore require simple tools for stability analysis that do not require large computational effort. Many indices have been developed by researchers studying voltage stability of distribution system as a measure of how far or near a system is from voltage instability or voltage collapse. Some of the common indices include Voltage Stability Margin (VSM) [4], real Power Stability Index (PSI) [9], Voltage Stability Index (VSI) [15]] and L-index [16].

The Voltage Stability Index (VSI) presented by Charkravorty and Das [14] will be used in this work as the objective function. The index is simple and suitable for voltage stability determination in radial distribution networks. The VSI is formulated by equation (1).

$$VSI_i = V_s^4 - 4V_s^2\left(R_iP_{Li} + X_iQ_{Li}\right) - 4\left(X_iP_{Li} - R_iQ_{Li}\right)^2 \quad (1)$$

where,

VSI_i is Voltage Stability Index at Bus i

V_s is distribution substation voltage which is 1 p.u

R_i is the resistance between source bus and bus i

X_i is reactance between source bus and bus i

P_{Li} is active power flow through bus i

Q_{Li} is reactive power flow through bus

B. *Search Heuristics*

For largest time of Power System Engineering, numerical optimization methods have been applied in power system planning and operation. Power system optimization is important since it has contributed to saving cost in terms of fuel cost, improved operational reliability and system security. Power systems have become large and complex and there has been a need to develop optimization techniques that will accommodate the large number of constraints in solving power system optimization problems. These methods are required to provide better solutions and have shorter computation times.

In recent times, researchers have widely used artificial intelligence methods for solving optimization problems because they can deal with complex problems that cannot be solved efficiently using conventional methods [17]. In addition, artificial intelligence methods have simple mathematical structure and simulate natural phenomena such as behavior of animals. Search heuristics are particularly applicable when objective functions are highly nonlinear and when the number of variables and constraints is large [18]. In addition, search heuristics reduce development time. This section gives a brief overview of Evolutionary Programming be used as the optimization method in this work.

Evolutionary Programming (EP) is a computational optimization method in the area of evolutionary computation which uses principles of natural evolution to find global optimal solution of a given problem [17]. As observed in [18], evolutionary programming is a useful method of optimization when other techniques such as gradient descent or direct analytical discovery methods are not possible and where the

search space may contain many local optimum solutions.

Evolution Programming is a general global optimization tool. EP seeks the optimal solutions over a number of generations or iterations. Evolution Programming uses mutation operator to generate a new population from an existing population. This technique is suitable in solving problems with real valued function optimization. The general procedure of Evolution Programming is as presented in fig. 3. EP has been applied by researchers to solve several power systems engineering problems. In [23], EP has been used to solve long term transmission network planning, estimating the transient and sub-transient parameters of a generator under normal operation, solving economic dispatch problem of units with non-smooth input-output characteristic functions and also solving the optimal power flow problem in FACTs.

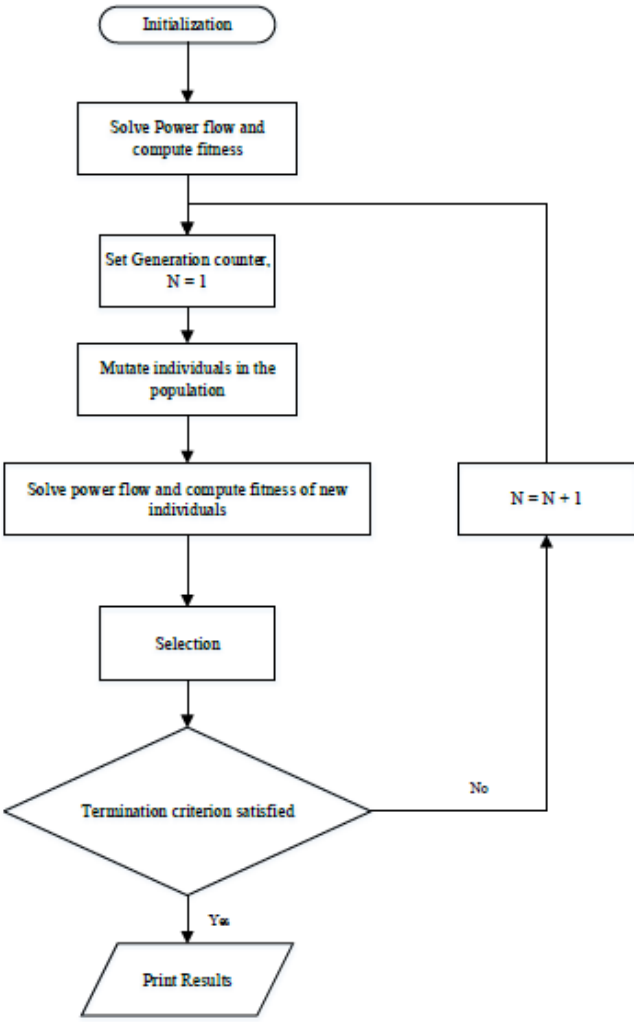


Fig. 3 General Evolution Programming Algorithm

III. METHODOLOGY

The objective function that was used is the Voltage Stability Index and the optimization problem was formulated as shown in the following section. This index varies from 0 to 1, with zero representing voltage collapse point and 1 representing the most stable bus. The bus at which voltage collapse is most probable was identified.

In order to identify buses most prone to voltage collapse, voltage stability indices for all buses were calculated and ranked from the smallest to the largest. The maximum number of DGs locations was determined by dividing the maximum DG penetration by the maximum DG size allowable. This number determined the number of buses corresponding to the lowest ranked indices that were candidates for DG and Capacitor placement.

The DGs were taken to operate in PQ mode with a power factor of 1. They therefore produce only active power at the buses they are installed at. Reactive power is provided by the capacitors which are simultaneously placed at the candidate buses together with the DGs.

Once the location for placement is identified, the next task was to identify the optimal size of the DGs and Capacitors that will improve the Voltage Stability Indices of the candidate buses without violating the system constraints. To identify the optimal size, Evolution Programming Algorithm was used to search for optimal sizes that maximize the voltage stability index. The test networks that will be used are IEEE-33 which is shown in fig. 4. The network data is shown in Appendix A. All simulations were done using MATLAB

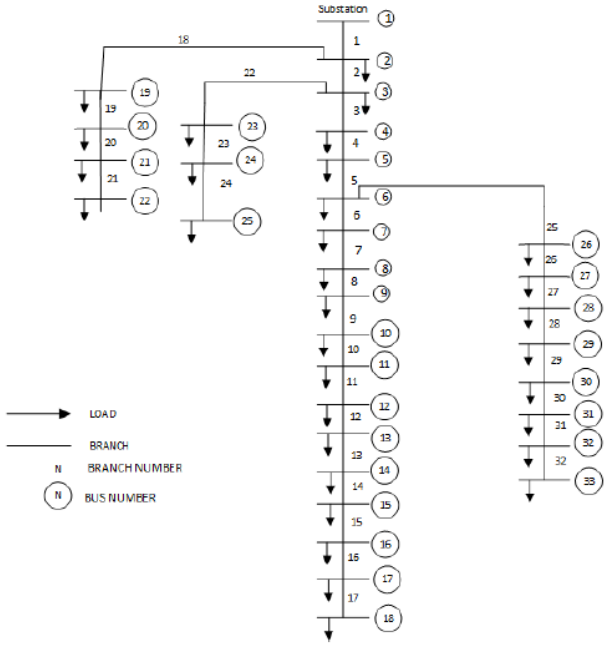


Fig. 4 Standard IEEE 33-bus system

A. *Problem formulation*

The optimization problem was formulated as follows:

$$\max(VSI) = \min(VSI_1, VSI_2, ...VSI_n) \quad (2)$$

Where

$$VSI_i = V_s^4 - 4V_s^2\left(R_i P_{Li} + X_i Q_{Li}\right) - 4\left(X_i P_{Li} - R_i Q_{Li}\right) \quad (3)$$

And $i = 2,3,4....n$,

$n = \text{no. of buses}$

Q_{Li} is the total reactive power fed through node i

P_{Li} is the total active power fed through node i

$i = 2,3,4....n$ R_i is the resistance between source and bus i

X_i is the reactance between source and bus i

For purposes of computation, equation (2) is modified as follows:

$$\max F(VSI) = \sum_{i=1}^{n} VSI_n \tag{4}$$

Equation (4) is maximized subject to the following constraints.

1) Equality constraints

The equality constraints include non-linear recursive power flow equations formulated as follows:

$$P_{i+1} = \left[P_{i,i+1} - \left(R_{i,i+1} \frac{P_{i,i+1}^2 + Q_{i,i+1}^2}{|V_i|^2} \right) - P_{i+1}^L + \alpha_{P_{DG}} P_{i+1}^{DG} \right] \tag{5}$$

$$Q_{i+1} = \left[Q_{i,i+1} - \left(X_{i,i+1} \frac{P_{i,i+1}^2 + Q_{i,i+1}^2}{|V_i|^2} \right) - Q_{i+1}^L + \alpha_{Q_C} Q_{i+1}^C \right] \tag{6}$$

and,

$$|V_{i+1}|^2 = |V_i|^2 - 2\left(P_{i,i+1}R_{i,i+1} + Q_{i,i+1}X_{i,i+1}\right) + \left(P_{i,i+1}^2 + Q_{i,i+1}^2\right)\frac{R_{i+1}^2 + X_{i+1}^2}{|V_i|^2} \tag{7}$$

where,

$i = 1,2,3,\mathrm{K}\ ,n$

P_{i+1} is active power flow through node $i+1$

Q_{i+1} is reactive power flow through node $i+1$

$|V_i|$ is the voltage magnitude at node i

$P_{i,i+1}$ is the active power flow in the branch between node i and $i+1$

$Q_{i,i+1}$ is the reactive power flow in the branch between node i and $i+1$

P_{i+1}^L is the active power load at node $i+1$

α_{DG} is the active power multiplier set to 1 where there is DG and 0 where there is none

Q_{i+1}^{C} is the capacitive reactive power load at node $i+1$

Q_{i+1}^L is the reactive power load at node $i+1$

α_{QC} is the reactive power multiplier set to 1 where there is a capacitor and 0 where there is none

$R_{i,i+1}$ is the resistace of branch between node i and $i+1$

$X_{i,i+1}$ is the reactance of branch between node i and $i+1$

2) Inequality constraints

a) Voltage operational tolerance

$$V_{i\min} \le V_i \le V_{i\max} \tag{8}$$

b) Thermal capacity limits

$$\left|I_{i,i+1}\right| \le \left|I_{i,i+1}\right|_{\max} \tag{9}$$

c) Total DG capacity constraint which should be within a given penetration level.

$$\frac{\sum_{i=1}^{n} P_{i+1}^{DG}}{P_{Load}} \le \eta \tag{10}$$

d) DG active power limits

$$P_{DG\min} \le P_i^{DG} \le P_{DG\max} \tag{11}$$

e) Capacitor reactive power limits

$$Q_{C\min} \le P_i^{DG} \le Q_{C\max} \tag{12}$$

where,

n – total number of buses in the distribution system

B. Parameter setting

The optimization problem parameters were set as shown in Table 1.

TABLE 1
OPTIMIZATION PROBLEM PARAMETERS

Parameter	Value
Maximum DG penetration, η	50%
Maximum DG Size	500kW
Maximum voltage at a bus $\|V_i\|_{\max}$	1.05p.u
Minimum voltage at a bus $\|V_i\|_{\min}$	0.95p.u
Total network active power demand	3.715MW
Maximum DG penetration in kW	1857.5kW
No. of locations=Max. DG penetration/max. DG size	4

C. Load flow

In order to calculate the initial network parameters, a load flow was done. The load flow technique used was the backward/forward sweep load flow method. In addition to the load flow, the voltage stability indices (VSI) for the network were calculated and a ranking done to determine the locations with lowest VSI that would be candidates for installation of DGs and capacitors. The load flow results and corresponding VSI are as shown in Table 2.

TABLE 2
LOAD FLOW DATA AND VSI FOR IEEE-33 BUS NETWORK

Bus No.	Vbus	VSI	Rank	Bus No.	Vbus	VSI	Rank
1	1	1	33	18	0.9036	0.669	1
2	0.997	0.9993	32	19	0.9965	0.938	27
3	0.983	0.9846	30	20	0.9929	0.986	31
4	0.975	0.9314	26	21	0.9922	0.972	29
5	0.968	0.9033	23	22	0.9915	0.969	28
6	0.949	0.8739	21	23	0.9792	0.915	24
7	0.946	0.8119	19	24	0.9725	0.919	25
8	0.932	0.7987	17	25	0.9692	0.894	22
9	0.926	0.7545	14	26	0.9474	0.817	20
10	0.92	0.7343	13	27	0.9448	0.805	18
11	0.919	0.7161	10	28	0.9334	0.796	16
12	0.918	0.7134	9	29	0.9251	0.759	15
13	0.911	0.7085	8	30	0.9216	0.732	12
14	0.909	0.6898	5	31	0.9174	0.721	11
15	0.908	0.6829	4	32	0.9164	0.708	7
16	0.906	0.6787	3	33	0.9161	0.705	6
17	0.904	0.6746	2				

From table 2, the buses that formed candidates for installation of DG and capacitor are 18, 17, 16 and 15 owing to their low VSIs.

D. Evolution programing Procedure

After identifying the location for installation of DGs and capacitors, Evolution Programming method was used to obtain the optimal sizes. The procedure was as shown in fig. 5.

IV. RESULTS AND DISCUSSION

After implementing the Evolutionary algorithm, the sizes of DGs and Capacitors at various buses were as shown in Table 3. In addition, the voltage magnitude and Voltage Stability indices after installation of DGs and capacitors were shown in Table 4. The voltage profile and voltage stability indices for all buses were plotted as shown in fig. 6 and 7 respectively.

TABLE 2
OPTIMAL SIZES OF DGS AND CAPACITORS

Bus No.	Size of DG(kW)	Size of Capacitor (kVAr)
15	80.3	77.6
16	74.4	39.2
17	126	117.3
18	214	217.5

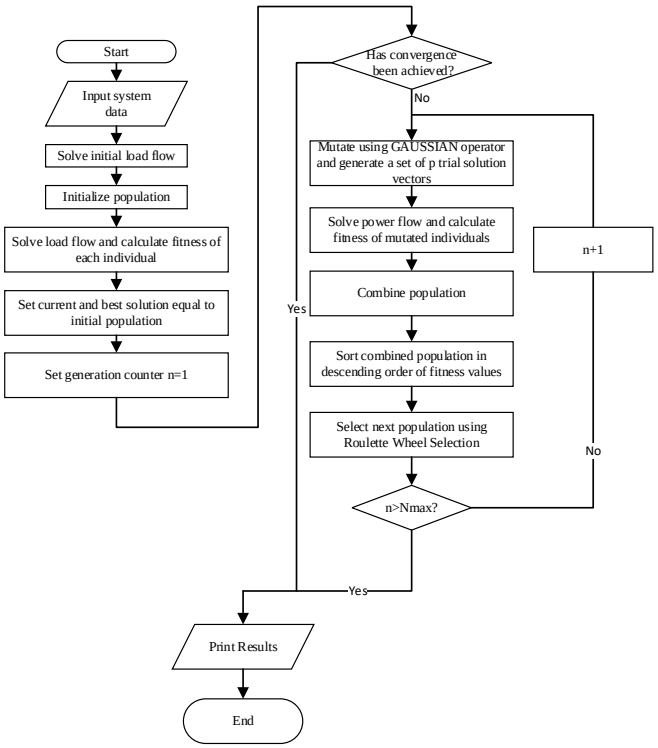


Fig. 5 Evolutionary Programming Algorithm

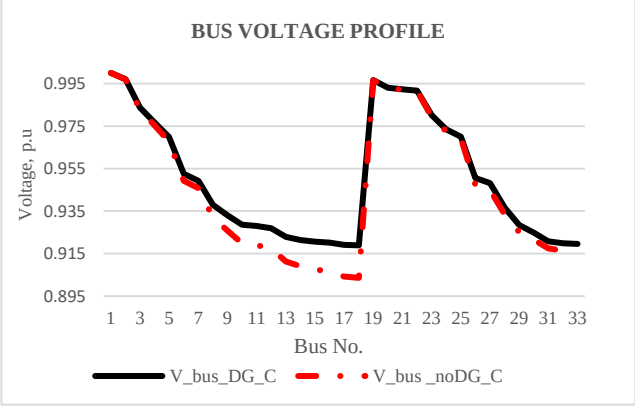


Fig. 6 Bus Voltage profile for IEEE 33 bus network

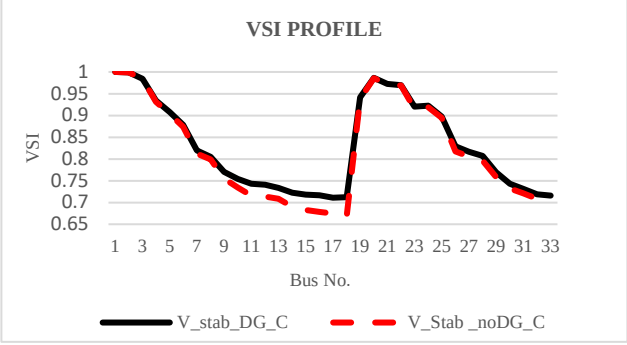


Fig. 7 VSI profile for IEEE 33 Bus network

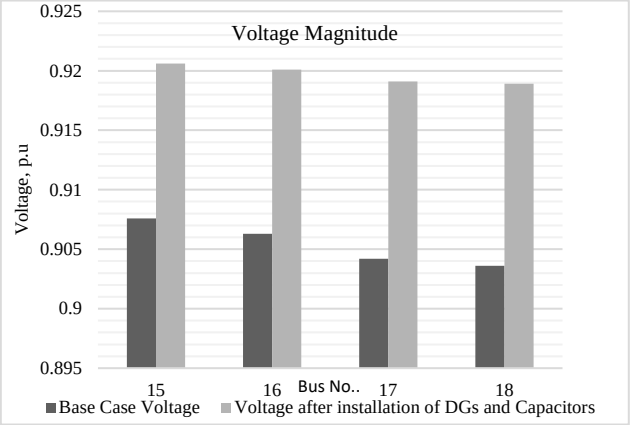


Fig. 8 Chart showing comparison of voltage magnitude at buses where DGs and Capacitors are installed

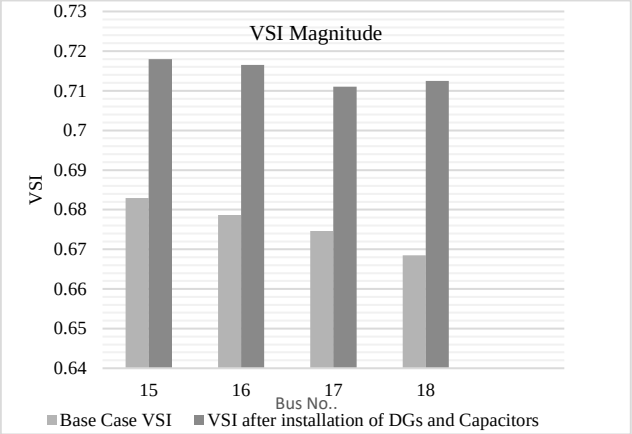


Fig. 9 Chart showing comparison of Voltage Stability Indices at buses where DGs and Capacitors are installed

Fig. 6 represents the voltage profile at all buses of the network before and after DG installation. The lowest value of voltage before installation of capacitors and DGs was 0.9036 p.u at bus 18 but after installation of DGs and capacitors the lowest value of voltage was 0.9189 p.u at bus 18 which represents 1.7% improvement. In addition, from fig. 7, it was noted that the lowest value of VSI before installing DGs and capacitors was 0.669 at bus 18. However, after installing DGs and capacitors, the lowest value of VSI became 0.7125. This represents an improvement of 6.6% of VSI on the weakest bus.

Fig. 8 and fig. 9 show a graphical representation of voltage and VSI values before and after installation of DGs and Capacitors. These are the buses where there is maximum impact on voltage profile and voltage stability index as a result of installing DGs and Capacitors. This can be attributed to the fact that voltage stability is a local phenomenon and not a global phenomenon. The maximum impact of voltage stability and voltage profile improvement will therefore be at the points where the DGs and capacitors will be installed. In other buses, the improvement will be minimal.

V. CONCLUSION

This work has demonstrated the application of search heuristics in optimizing the voltage stability of radial distribution network through installation of optimal sizes of distributed generation and capacitors. In addition, this work has introduced a novel method of determining the best location for placing distributed generation and capacitors in a given radial distribution network.

Through this technique, it has been demonstrated that distribution voltage profile and voltage stability can be greatly improved by optimal placement of distributed generation and capacitors.

APPENDIX

TABLE 3:
IEEE 33 BUS NETWORK DATA

Sending Bus	Receiving Bus	R(Ohms)	X(Ohms)	Load at Receiving Bus P	Q
1	2	0.0922	0.0477	100	60
2	3	0.4930	0.2511	90	40
3	4	0.3660	0.1864	120	80
4	5	0.3811	0.1941	60	30
5	6	0.8190	0.7070	60	20
6	7	0.1872	0.6188	200	100
7	8	1.7114	1.2351	200	100
8	9	1.0300	0.7400	60	20
9	10	1.0400	0.7400	60	20
10	11	0.1966	0.6550	45	30
11	12	0.3744	0.1238	60	35
12	13	1.4680	1.1550	60	35
13	14	0.5416	0.7129	120	80
14	15	0.5910	0.5260	60	10
15	16	0.7643	0.5450	60	20
16	17	1.2890	1.7210	60	20
17	18	0.7320	0.5740	90	40
2	19	0.1640	0.1565	90	40
19	20	1.5042	1.3554	90	40
20	21	0.4095	0.4784	90	40
21	22	0.7089	0.9373	90	40
3	23	0.4512	0.3083	90	40
23	24	0.8980	0.7091	90	40
24	25	0.6960	0.7011	520	200
6	26	0.2030	0.1034	320	200
26	27	0.2842	0.1447	60	25
27	28	1.0590	0.9337	60	20
28	29	0.8042	0.7006	120	70
29	30	0.5075	0.2585	200	600
30	31	0.9744	0.9630	250	70
31	32	0.3105	0.3619	210	100
32	33	0.3410	0.5302	60	40

System voltage is 12.66kV

REFERENCES

[1] T. Hashim, M. A. and H. Shareef, "A Review on Voltage Control Methods for Active Distribution Systems," PRZEGLAD ELEKTROTECHNICZNY (Electrical Review), vol. 88, no. 6, pp. 304-312, 2012.

[2] T. Xu and P. Taylor, "Voltage Control Techniques for Electrical Distribution Networks including Distributed Generation," in 17th World Congress, The International Federation of Automatic Control, Seoul, 2008.

[3] X. Hu, X. Zhang, X. Lou, X. Men, J. L. Zhang and S. Pan, "Analysis on the Voltage Stability and the Allowed Penetration Capacity of Distributed Generations in Yuhang Power Grid," in 9th International Conference on Power Electronics-ECCE Asia, Seoul, 2015.

[4] A. Abri, F. El-Saadany and A. M., "Distributed Generation Placement and Sizing Method to Improve the Voltage Stability Margin in a Distribution System," IEEE Transactions on Power Systems, vol. 28, pp. 326-334, 2013.

[5] K. Sarmin, W. Nakawiro, M. Wani, F. Siam, Z. Hussien, A. Ibrahim and A. Hussin, "Co-ordinated Voltage Control in Distributin Network with Renewable Energy Based Distributed Generation," Scientific Research, vol. 5, no. 1, pp. 208-214, 2013.

[6] P. Georgilakis and N. Hatziargyriou, "Optimal Distributed Generation Placement in Power Distribution Networks: Models, Methods and Future Research," IEEE Transactions on Power Systems, vol. 28, pp. 3420-3428, 2013.

[7] L. Wong, H. Shareef, A. Mohamed and A. Ibrahim, "Optimal Battery Sizing in Photovoltaic Based Distributed Generation using Enhanced Opposition-Based Firefly Algorithm," The Scientific World Journal, 2014.

[8] S. Neenu and R. Antony, "Optimal DG Placement Considering Voltage Stability Enhanccement using PSO," in 2013 International Conference on Control Communication and Computing (ICCC), Thiruvananthapuram, 2013.

[9] J. Jamian, M. Aman, M. Mustafa, G. Jasmon, H. Mokhlis, A. Bakar and M. Abdullah, "Optimum Multi DG units Placement and Sizing Based on Voltage Stability Index and PSO," in 47th International Universities Power Engineering Conference, London, 2012.

[10] M. Imran and M. Kowsalya, "Optimal Distributed Generation and Capacitor placement in Power Distribution Networks for Power Loss Minimization," in Advances in Electrical Engineering(ICAEE), Vellore, 2014.

[11] A. Mohapatra, S. Behera, S. Nayak and B. Panigrahi, "A Study on DG and Capacitor Placement in Radial Distribution System," in IEEE International Conference on Power Elecronics , Bengaluru, 2012.

[12] M. Kasaei, "Optimal Placement of Distributed Generation and Capacitor in Distribution Networks by Ant Colony Algorithm," International Journal on Technical and Physical Problems of Engineering, vol. 6, no. 20, pp. 52-56, 2014.

[13] F. Viawan, "Voltage Control and Voltage Stability of Power Distribution Systems in the Presence of Distribution Generation," PhD thesis, Chalmers University of Technology, Goteborg, 2008.

[14] M. Alonso and H. Amaris, "Voltage Stability in Distribution Network with DG," in 2009 IEEE Bucharest Power Tech Conference, Bucharest, 2009.

[15] M. Chakravorty and D. Das, "Voltage Stability Analysis of Radial Distribution Networks," Electrical Power and Energy Systems, vol. 23, no. 2, pp. 129-135, 2001.

[16] P. Balachennaiah, P. Reddy and U. Raju, "A Novel Algorithm for Voltage Stability Augmentation through Optimal Placement and Sizing of SVC," in 2015 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), Kozhikode, 2015.

[17] W. Ongsakul and D. Vo, Artificial intelligence in Power Systems Optimization, Florida: Taylor & Francis Group, LLC, 2013.

[18] K. Lee and M. El-Sharkawi, Modern Heuristic Optimization Techniques: Theory and Applications to Power Systems, Hoboken, New Jersey: John Wiley & Sons,Inc, 2008.

Voltage Stability Analysis of Power System with Large Wind Power Integration

Weldon K. Koskei [1], David K. Murage[2], Samuel Kangethe[3] and Michael J. Saulo[4]

Abstract—Renewable energy deployment is currently advocated globally as a key component in addressing current climate change, providing access to green energy and increasing energy services as well as creating economic opportunities. Wind power generation is one of renewable energy sources which is gaining more interest as a generation source to add more power capacity to power system grid. This is due to the fluctuation of fuel prices affecting thermal power and unfavorable weather conditions affecting hydropower. According to world wind energy association, the total world wind power installed capacity had reached 372GW by end of 2014 and is estimated to reach 8% of world electricity market by 2035. Many governments in the world today have enacted policies to promote renewable energy deployment to meet the increasing electricity demand and reduce dependence of power from hydro and thermal plants.

Wind induction generators are non-conventional and use different technology and their output depend upon changing weather conditions. This generate power quality challenges such as uncontrollable reactive power and voltage fluctuations which deteriorate power system voltage stability. Since the rate of integrating wind power generation to transmission power systems is increasing rapidly, there is need to investigate their impact on power system voltage stability. This in turn will help power system planners and operators in their daily power system operations. This paper aims at investigating the impact of wind power integration on power system voltage stability. Fast Voltage Stability Index (FVSI) and Line stability Index (L_{mn}) were used. These line voltage indices indicate the proximity of transmission lines to voltage instability. IEEE 30 bus system was used as a test system to carry out this research. The research also investigated impacts of different location and different penetration levels of wind power into the power system using MATLAB simulation software.

Keywords— Induction Generators, Voltage Stability, Fast Voltage Stability Index, Line Stability Index.

I. INTRODUCTION

Wind energy is currently playing an important role in establishing a friendly and environmentally sustainable economy which is low in carbon. It was initially used as a source of energy to drive ships, grind grains and pump water. Wind energy was first used to generate electricity in the late 19th century and it gained more interest in around 1970s due to oil crisis and more awareness on conservation of resources [1]. Wind energy was later used in remote power systems, charging batteries, isolated or island power systems and in residential power systems. The current available wind turbines ranges from 100kW to 5MW. The wind power generation rapidly increased due to industrial and governmental initiatives.

The wind turbines development have increased rapidly which has led to continuous growth of global wind market. Most of the electricity industries globally are currently not able to meet power demand for domestic and industrial purposes. Therefore there is need for more electricity generation so that it can provide electricity to larger population as well as supporting economic growth. Electric power generated from hydropower is not reliable during dry seasons and thermal power plants uses oil products which is greatly affected by ever rising oil prices leading to negative impacts on economic activities. Most governments in the world have enacted policies to promote renewable energy deployment to meet the increasing electricity demand and reduce dependence of power from hydro and thermal plants. Wind power generation as one of the renewable energy is gaining more interest in adding more power generation capacity to power system grid [2], [3]. The nature of wind power generation poses integration stability challenges to the existing electric power systems.

The generator technology used in most wind turbines is induction generator technology which consume reactive power from the grid. This wind generator technology together with variation in wind energy generation lead to uncontrollable reactive power, voltage distortion, voltage fluctuations as well as power fluctuations which affect power quality and stability of the power system.

The high level of industrialization globally has led to increase demand for electric power but the expansion of power system infrastructure is limited by financial constraints. This has led to operation of power system close to its operating stability limits. This overloaded operation condition of power system leads to instability challenges like voltage instability due to inability of the power system to supply required reactive power. Integration of induction wind generators which consume reactive power from already stressful power system might worsen the condition.

The impacts of wind power integration on voltage stability of power system was investigated and results obtained from simulating standard IEEE 30 bus system were discussed.

II. VOLTAGE STABILITY

The demand for electric power is ever increasing in many power systems globally due to rapid growth in industrialization.

Weldon K. Koskei, MSc Student, Department of Electrical and Electronic Engineering, JKUAT, (phone: +254729918244; e-mail: *wkoskei@jkuat.ac.ke*),
David K. Murage, Department of Electrical and Electronic Engineering, JKUAT, (e-mail: *dkmurage25@yahoo.com*),
Samuel Kangethe, Department of Electrical and Electronic Engineering, JKUAT, (e-mail: samuel.kangethe@jkuat.ac.ke),
Michael J. Saulo, Department of Electrical and Electronic Engineering, Technical University of Mombasa, TUM, (e-mail: michaelsaulo@tum.ac.ke),

This has led to operation of power systems closer to stability limits which make the system vulnerable to instability challenges like voltage instability. Many researchers have developed interest in studying voltage stability due to its cause for blackout incidents in many electric markets globally.

Voltage stability is referred to as ability of the power system to maintain steady voltages at all buses in the system under a normal operating conditions and after being subjected to a disturbance [4], [5]. Voltage instability is mainly caused by increase in the load demand and inability of the system to supply required reactive power which leads to uncontrollable voltage drops. Voltage stability is classified into large disturbance and small disturbance voltage stability. Large disturbance voltage stability is concerned with maintaining voltages after large disturbances such as faults and loss of generation in the system while small disturbance voltage stability deals with maintaining voltages after small disturbances such as gradual changes of the loads in the system.

There are tools used to approximate the voltage stability status of the power system. Voltage stability indices are examples of these tools used in power system to gauge the proximity of a particular operating point to voltage instability state. They shows how close a given operating point is to the steady state voltage stability margin. They are helpful to power system operators in their system operations since they show a real time state of the power system. They are also useful during power system designing and planning of operations. Fast voltage stability index developed by [6] and line stability index developed by [7] were used in this paper to show how close the system is to voltage collapse. The line stability index (L_{mn}) is formulated as follows:

$$L_{mn} = \frac{4X_{ij}Q_j}{[V_i \mathrm{Sin}(\theta-\delta)]]^2} \quad (1)$$

Where

X_{ij} - line reactance
Q_j - receiving end reactive power
V_i - sending end voltage
θ - line impedance angle
δ - angle between sending end and receiving end voltage

The value of L_{mn} is used to show how close the line is to voltage collapse. Lines with L_{mn} values close to zero are considered stable but those lines with values close to 1 are considered unstable. This index is capable of identifying critical areas and most critical lines in a power system network. This Fast Voltage stability index ($FVSI_{ij}$) is formulated as:

$$FVSI_{ij} = \frac{4Z_{ij}^2 Q_j}{V_i{}^2 X_{ij}} \quad (2)$$

Where

Z_{ij} - line impedance
Q_j - receiving end reactive power
V_i - sending end voltage
X_{ij} - line reactance

This index shows the proximity of a line to voltage collapse. The lines with FVSI value close to 0 are considered stable while those with values close to 1 are termed as unstable. Just like L_{mn} index, FVSI index have capability of identifying critical lines and most critical areas in a power system network.

III. TYPES OF WIND POWER GENERATORS

Wind generator technology is one of the limiting factor in development of wind turbines.

AC Asynchronous WTGs have been widely used in wind turbine applications and below are two main types:

A. Fixed speed induction generators (FSIG)
B. Doubly fed induction generators (DFIG)

A. Fixed Speed Induction Generators (FSIG)

FSIGs have squirrel cage rotors and normally referred to as squirrel cage induction generators SCIGs [8]. Fig 1 shows schematic diagram of FSIG

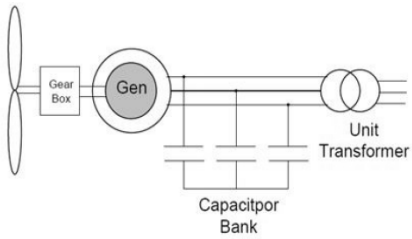


Fig 1: Schematic Representation of FSIG

The stator is connected to the power system grid through a transformer while the rotor is connected to the wind turbine system through a gearbox. The rotor is considered to run within a narrow range of speeds hence termed as fixed speed induction generators. FSIGs normally operate at 1500 rpm for 50Hz electric power system grid [9].

FSIGs do not have capability to control output voltage and hence reactive power are supplied from external sources. These WTGs are also large in size, noisy and have low efficiency.

B. Doubly Fed Induction Generators (DFIG)

DFIGs have wound rotors and are currently widely used in wind power generation [10]-[12]. The commercial wind turbine systems of DFIGs have increased to 5MW. Fig 2 shows schematic diagram of DFIG.

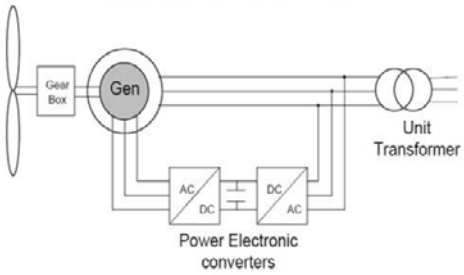


Fig 2: Schematic Representation of DFIG

The stator is connected to the power system network through transformers while the rotor is connected to the grid through PMW power converters. These power converters control rotor circuit current, frequency as well as phase angle shifts. They operate at wide range of speeds hence can produce high energy,

reduced mechanical stress and power fluctuations as well as providing capability for reactive power control.

DFIGs have capability to operate in sub-synchronous and super-synchronous states. The stator transfers power to the power system grid only while the rotor can transfer power in both directions with the help of PMW power converters which can supply voltage and current at different phase angles. During sub-synchronous, rotor side converter (RSC) acts as an inverter while grid side converter (GSC) act as a rectifier and active power flows from the grid to the rotor. In super-synchronous state, RSC act as rectifier while GSC operates as an inverter and hence active power will flow from stator and rotor to the power system grid.

IV. TEST SYSTEM AND SIMULATIONS RESULTS

The impact of wind power integration on power system voltage stability was investigated by performing load flow analysis on standard IEEE 30 bus system and its system data is as shown in the appendix. This test system is as shown in Fig 3. This system comprises of six generator buses, twenty four load buses and forty one interconnected transmission lines. Static loads were considered in this simulations.

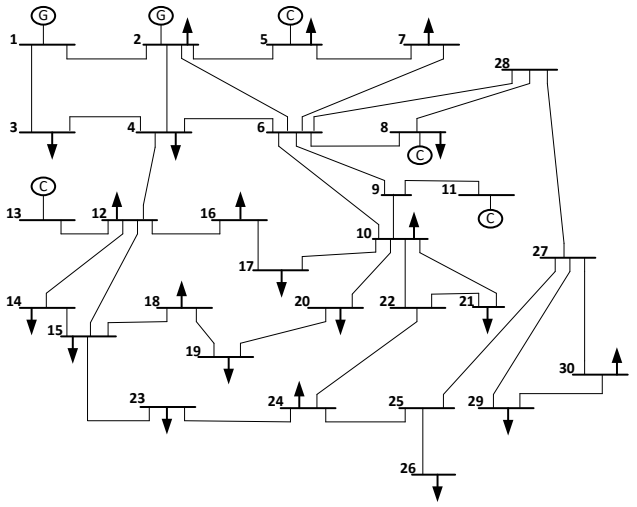


Fig 3: Standard IEEE 30 Bus Test System

Double fed induction generators were used to carry out this study. The simulations were carried out first without wind power integration and after wind power integration into test power system.

Different wind power integration locations were considered and in each location different penetration levels were analyzed from 10% to 30% of the total power generated. The results obtained in each case were used to calculate the voltage stability indices of some lines and these indices compared to quantify how the inclusion of wind power integration affects voltage profiles of the power system. Table I shows calculated indices on some lines when wind power was integrated at bus 5 and table II shows results obtained when point of integration is bus 15 while table III results obtained when bus 25 considered as the integration point. All simulations in this study were carried out at base loading.

Table I: Wind power integration at bus 5

From Bus	To Bus	0%		10%		20%		30%	
		FVSI	L_{mn}	FVSI	L_{mn}	FVSI	L_{mn}	FVSI	L_{mn}
2	5	0.014	0.012	0.014	0.012	0.014	0.013	0.015	0.014
5	7	0.020	0.018	0.020	0.018	0.021	0.020	0.025	0.022
2	4	0.021	0.023	0.021	0.023	0.023	0.024	0.030	0.028
2	6	0.024	0.022	0.024	0.022	0.025	0.024	0.039	0.037
1	2	0.009	0.007	0.009	0.007	0.010	0.008	0.023	0.019
6	7	0.025	0.023	0.025	0.023	0.027	0.025	0.049	0.043
16	17	0.016	0.015	0.016	0.015	0.016	0.015	0.016	0.025
27	29	0.253	0.257	0.253	0.257	0.253	0.257	0.253	0.257

The values of the indices calculated for some branches in table I above were obtained by first carrying out simulations for base loading without wind power integration. The wind power was introduced into the system first at 10% of the total generated power and then 20% and finally 30%. The voltage stability indices were calculated for each integration case and compared with those obtained without wind power integration into the system. Results in table I shows that there is little impact when wind power integrated at 10% but there is slight impact at 20% integration. There is also small decrease of voltage profiles at 30% wind power integration shown by higher values of indices as compared with other cases. Bus 5 which is a point of wind power integration in this case is considered as generation area since most generators are concentrated around this bus. Hence the reactive power drawn by induction wind turbines generators is easily compensated leading to small impact on voltage stability. It is also noted that there is no impact on the lines far away from the point of integration.

Table II: Wind power integration at bus 15

From Bus	To Bus	0%		10%		20%		30%	
		FVSI	L_{mn}	FVSI	L_{mn}	FVSI	L_{mn}	FVSI	L_{mn}
14	15	0.012	0.011	0.053	0.053	0.143	0.129	0.278	0.262
12	15	0.032	0.030	0.060	0.049	0.120	0.122	0.238	0.228
15	18	0.021	0.020	0.062	0.058	0.136	0.135	0.263	0.263
15	23	0.035	0.038	0.062	0.061	0.139	0.137	0.273	0.273
12	14	0.029	0.027	0.030	0.029	0.060	0.043	0.101	0.082
18	19	0.037	0.035	0.043	0.043	0.072	0.069	0.124	0.119
2	5	0.014	0.013	0.014	0.013	0.014	0.012	0.014	0.012
27	29	0.253	0.257	0.253	0.257	0.267	0.269	0.293	0.282

The same procedure was repeated by first carrying out load flows simulations without wind power integration and then integrate it at bus 15 which is considered as transmission area. The branches around the point of integration were mostly analyzed to assess the impact of wind power inclusion. Just like the first case, there is small impact at 10% penetration level but higher effect at 30% as shown by higher indices values. There is higher decrease of voltage stability at this point of integration as compared with the first case as shown by higher values of the indices. This is because at this integration point there is less

reactive power as compared with generation area in the first case.

Table III: Wind power integration at bus 25

From Bus	To Bus	0%		10%		20%		30%	
		FVSI	L_{mn}	FVSI	L_{mn}	FVSI	L_{mn}	FVSI	L_{mn}
25	26	0.254	0.252	0.280	0.262	0.302	0.281	0.382	0.360
24	25	0.047	0.048	0.241	0.241	0.290	0.288	0.311	0.302
25	27	0.031	0.029	0.258	0.249	0.313	0.302	0.383	0.384
22	24	0.023	0.022	0.140	0.139	0.148	0.141	0.211	0.207
23	24	0.040	0.038	0.130	0.130	0.140	0.135	0.163	0.160
27	29	0.253	0.257	0.266	0.261	0.280	0.272	0.341	0.344
2	5	0.014	0.012	0.014	0.012	0.014	0.012	0.014	0.012
16	17	0.016	0.025	0.029	0.028	0.030	0.029	0.034	0.033

The results in table III were obtained by carrying out simulations at different wind power penetration at bus 25 which is considered as load area. Just like the other scenarios, the values of the indices generally increases with increase in penetration level.

There is higher decrease in voltage stability as indicated by higher indices when the point of wind power integration is electrically far away from generation area as shown in table III. The simulations without wind power integration shows that the values of indices at load area are higher than those at generation or transmission area. This indicate that buses at load area are weak as compared to those at generation and transmission area. Hence inclusion of induction wind power generators at load area further decrease voltage profiles and impact increases with the increase in penetration levels.

V. CONCLUSION

The impact of wind power integration on power system voltage stability was investigated in this paper. The study shows that the impact vary with penetration level and integration point of the wind farms. There is more decrease of power system voltage stability at load area as compared to generation area. The impact also increase with the increase in penetration level. Therefore reactive power compensation will depend on wind power penetration level and the point of integration to the power system.

REFERENCES

[1] Manwell, J. F., Mc Gowan, J. G., & Rogers, A. L. (2009). Wind energy explained: theory, design and application. *2nd edition, Wiley, England*.

[2] Shanker, T., & Singh, R. K. (2012). Wind energy conversion system: A review. *Students Conference on Engineering and Systems (SCES)*, 1-6.

[3] World wind energy association, Quarterly Bulletin, Special Issue 2015.

[4] P. Kundur, Power System Stability and Control, McGraw-Hill, 1993

[5] C. W. Taylor, Power System Voltage Stability McGraw-Hill, 1994.

[6] T. K. A. R. Ismail Musirin, "Novel Fast Voltage Stability Index (FVSI) for Voltage Stability Analysis in Power Transmission System," in *Student Conference on Research and Development Proceeding,*, 2002, pp. 265–268.

[7] M. Moghavvemi, "New method for indicating voltage stability in power system" Proceedings of IEEE International Conference on Power Engineering, Singapore, lPEC, 1997, pp. 223-227.

[8] Zhang, M., Wang, W., Chen, Y. R., & Coombs, T. (2012). Design methodology of HTSbulk machine for direct-driven wind generation. *IEEE Transactions on Applied Superconductivity*,22(3), 5201804.

[9] Abo-Khalil, A. G. (2011). A new wind turbine simulator using a squirrel-cage motor for wind power generation systems. *IEEE Ninth International Conference on Power Electronics and Drive Systems (PEDS)*, 750-755.

[10] Polinder, H. (2011). Overview and trends in wind turbine generator systems",. *IEEE Power and Energy Society General Meeting*, 1-8.

[11] Delli, Colli. V., Marignetti, F., & Attaianese, C. (2012). Analytical and multiphysics approach to the optimal design of a 10-MW DFIG for direct-drive wind turbines. *IEEE Transactions on Industrial Electronics*, 59(7), 2791-2799.

[12] Ruviaro, M., Runcos, F., Sadowski, N., & Borges, I. M. (2012). Analysis and test results of a brushless doubly fed induction machine with rotary transformer. *IEEE Transactions on Industrial Electronics*, 59(6), 2670-2677.

APPENDIX

IEEE 30 BUS SYSTEM DATA

Table A1 : Bus Data

Bus No	type	Pd	Qd	Vm	Va	Vmax	Vmin
1	3	0	0	1.06	0	1.06	0.94
2	2	21.7	12.7	1.043	-5.48	1.06	0.94
3	1	2.4	1.2	1.021	-7.96	1.06	0.94
4	1	7.6	1.6	1.012	-9.62	1.06	0.94
5	2	94.2	19	1.01	-14.37	1.06	0.94
6	1	0	0	1.01	-11.34	1.06	0.94
7	1	22.8	10.9	1.002	-13.12	1.06	0.94
8	2	30	30	1.01	-12.1	1.06	0.94
9	1	0	0	1.051	-14.38	1.06	0.94
10	1	5.8	2	1.045	-15.97	1.06	0.94
11	2	0	0	1.082	-14.39	1.06	0.94
12	1	11.2	7.5	1.057	-15.24	1.06	0.94
13	2	0	0	1.071	-15.24	1.06	0.94
14	1	6.2	1.6	1.042	-16.13	1.06	0.94
15	1	8.2	2.5	1.038	-16.22	1.06	0.94
16	1	3.5	1.8	1.045	-15.83	1.06	0.94
17	1	9	5.8	1.04	-16.14	1.06	0.94
18	1	3.2	0.9	1.028	-16.82	1.06	0.94
19	1	9.5	3.4	1.026	-17	1.06	0.94
20	1	2.2	0.7	1.03	-16.8	1.06	0.94
21	1	17.5	11.2	1.033	-16.42	1.06	0.94
22	1	0	0	1.033	-16.41	1.06	0.94
23	1	3.2	1.6	1.027	-16.61	1.06	0.94
24	1	8.7	6.7	1.021	-16.78	1.06	0.94
25	1	0	0	1.017	-16.35	1.06	0.94
26	1	3.5	2.3	1	-16.77	1.06	0.94
27	1	0	0	1.023	-15.82	1.06	0.94
28	1	0	0	1.007	-11.97	1.06	0.94
29	1	2.4	0.9	1.003	-17.06	1.06	0.94
30	1	10.6	1.9	0.992	-17.94	1.06	0.94

Table A2: Generator Data

Bus No	Pg	Qg	Qmax	Qmin	Vg	mBase	Pmax	Pmin
1	260.2	-16.1	10	0	1.06	100	360.2	0
2	40	50	50	-40	1.045	100	140	0
5	0	37	40	-40	1.01	100	100	0
8	0	37.3	40	-10	1.01	100	100	0
11	0	16.2	24	-6	1.082	100	100	0
13	0	10.6	24	-6	1.071	100	100	0

Table A3: Line Data

Line No	From bus	To bus	r	x	b	ratio
1	1	2	0.0192	0.0575	0.0528	0
2	1	3	0.0452	0.1652	0.0408	0
3	2	4	0.057	0.1737	0.0368	0
4	3	4	0.0132	0.0379	0.0084	0
5	2	5	0.0472	0.1983	0.0418	0
6	2	6	0.0581	0.1763	0.0374	0
7	4	6	0.0119	0.0414	0.009	0
8	5	7	0.046	0.116	0.0204	0
9	6	7	0.0267	0.082	0.017	0
10	6	8	0.012	0.042	0.009	0
11	6	9	0	0.208	0	0.978
12	6	10	0	0.556	0	0.969
13	9	11	0	0.208	0	0
14	9	10	0	0.11	0	0
15	4	12	0	0.256	0	0.932
16	12	13	0	0.14	0	0
17	12	14	0.1231	0.2559	0	0
18	12	15	0.0662	0.1304	0	0
19	12	16	0.0945	0.1987	0	0
20	14	15	0.221	0.1997	0	0
21	16	17	0.0524	0.1923	0	0
22	15	18	0.1073	0.2185	0	0
23	18	19	0.0639	0.1292	0	0
24	19	20	0.034	0.068	0	0
25	10	20	0.0936	0.209	0	0
26	10	17	0.0324	0.0845	0	0
27	10	21	0.0348	0.0749	0	0
28	10	22	0.0727	0.1499	0	0
29	21	22	0.0116	0.0236	0	0
30	15	23	0.1	0.202	0	0
31	22	24	0.115	0.179	0	0
32	23	24	0.132	0.27	0	0
33	24	25	0.1885	0.3292	0	0
34	25	26	0.2544	0.38	0	0
35	25	27	0.1093	0.2087	0	0
36	28	27	0	0.396	0	0.968
37	27	29	0.2198	0.4153	0	0
38	27	30	0.3202	0.6027	0	0
39	29	30	0.2399	0.4533	0	0
40	8	28	0.0636	0.2	0.0428	0
41	6	28	0.0169	0.0599	0.013	0

Grid Integration of Large Capacity Wind Power: A Review.

Joseph N. Mathenge [1], D.K Murage [2], J.N Nderu [3] and C.M Muriithi [4]

***Abstract*—** Since the adoption of the Kyoto Protocol in 1997, the world has inclined towards the integration of renewable energy sources into the grid with the aim of reducing carbon emissions in the environment. Developed and developing countries have continued investing millions of dollars into research and deployment of renewables and this has sparked a huge interest in this field by researchers aiming to streamline and make this technology efficient. Initially, renewables were adopted in small quantities posing negligible threat to the security and running of power systems. However, over the last two decades, there has been a huge interest in large capacity integration of renewable energy sources. The disadvantage of this is that some of the renewables have a stochastic nature which makes their integration pose challenges to the wider grid. These stochastic renewables include solar and wind. Between the two, wind is the most intermittent yet the most widely adopted renewable. This paper reviews the integration of large scale wind power – the present, the future and challenges being faced.

***Keywords*—**Large capacity, Renewable Energy, Wind Power.

I. Introduction

RENEWABLE energy sources are the interest of most if not all power systems across the world. The fact that they occur freely in nature gives them an edge over other sources since they do not have any running costs associated with them. The prices of fossil fuels are subject to forces of supply and demand as well as world politics. This makes the per-unit cost of power from fossil fuels volatile in nature [1]. Renewable energy sources incur a zero fuel cost and hence make them cheaper than fossil fuels. Other than the capital cost of installation of the technology, there are minimal running costs associated making them an attractive option.

Wind and solar have one disadvantage associated with them – they occur intermittently in nature. This natural variability makes it difficult to integrate the two sources with the grid and is referred to as non-dispatchability. Despite the fact that solar and wind have both seen considerable interest from researchers and power system operators, it is wind that has been mostly deployed in large capacity plants. Wind technology has matured to great extents in the last two decades and as the wind power plants continue to be deployed in present day grids, the issue of power system security becomes one of great concern that requires deep planning and evaluation [2].

This, however, has not discouraged the use of renewables in the grid. Despite the fact that China has substantial fossil fuel resources, it comes out as the country with the most ambitious wind and solar power targets in the world. China aims to have between 150GW -180GW of installed wind capacity by 2020. Developing countries have not been left behind either in this expansion and have been observed to be expanding their renewable energy demand at a rate of 3.5% per annum since 2009 [1]. For renewable energy to achieve better penetration and be adopted by many power systems, the government need to play a pivotal role in ensuring the adoption of this technology [3]. This happens in the form of legislation, encouraging feed in tariffs, tax credits and the development of smart grids. The result of adoption of renewable energy sources results in cheaper power tariffs, reduction of the global carbon footprint associated with power generation, and meeting the growing energy demand in the world among other benefits.

II. REVIEW

A. Wind Power Generation

Wing turbine generators harness wind power and convert it to electricity using an aerodynamic rotor. Wind acts as the prime-mover in this system of generation. Figure.1. below shows the structure of a wind turbine generator:

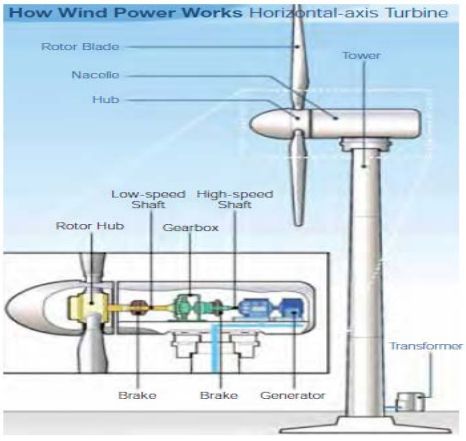


Fig. 1. A wind turbine generator [1].

Wind energy can be harnessed in two categories:

i. Onshore
ii. Offshore

Joseph N. Mathenge, Department of Electrical Engineering, JKUAT (Phone: +254725096991; e-mail: jmathenge@ jkuat.ac.ke).

Prof D.K. Murage, Department of Electrical Engineering, JKUAT
Prof J.N Nderu, Department of Electrical Engineering, JKUAT
Prof C.M Muriithi, Department of Electrical Engineering, MUT

Wind turbine technologies have seen the development of turbines with a capacity of 1.5MW – 3 MW for the onshore harnessing and 2MW – 5 MW for offshore. Larger versions of wind turbines are in the range of 5MW to 6MW. Offshore winds are more uniform and stronger than the onshore winds and hence more power can be obtained from them. However, offshore wind power harnessing involves a more complex installation system and thus not much work has been done on it. Typically, a wind power curve has three speeds: The cut in speed, the rated speed and a cut out speed [4]. The turbine begins working at the cut in speed, achieves maximum power output at the rated speed and stops working at the cut out speed to prevent the turbine from getting damaged [5].

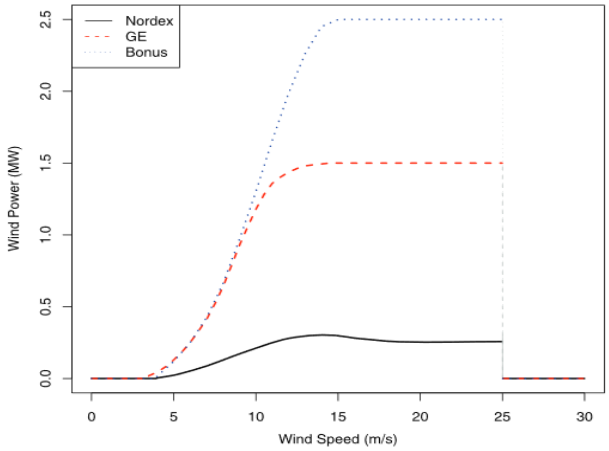


Fig. 2. Deterministic Power Curve Showing Cut-in Speed, Rated Speed and Cut-out Speed for Turbines with Capacities 0.3MW-Nordex, 1.5MW-GE and 2.5MW-Bonus.

Single wind turbine generators are limited in capacity and hence they are aggregated to form wind farms whose output now becomes considerable compared to the other sources in the grid. Their power is collated and injected into the grid as power from a single source [1].

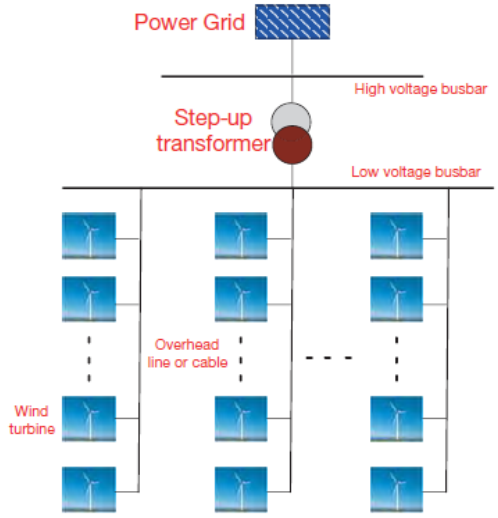


Fig. 3. Structure of a wind farm [1].

B. Wind Energy Conversion Systems (WECS)

There are four WECS: Type I, II, III and IV [5]

Type I: Fixed Speed Induction Generator:
It is based on the squirrel cage induction generator and as its name suggests its speed is fixed with little variations of around 1-2% [6] which means that its output power fluctuates as the speed of wind varies. Induction generators draw a lot of reactive power and hence the Type I wind energy generator is fitted with a reactive power compensation mechanism.

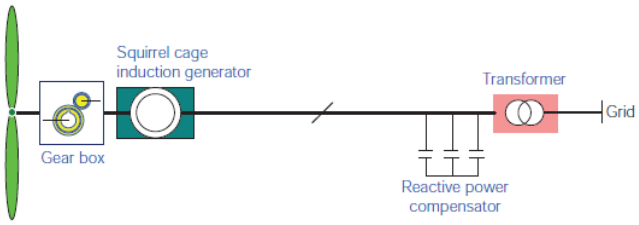


Fig. 4. Type I Wind Turbine Generator [1]

Type II: Variable Slip Induction Generator
It is based on the wound rotor induction generator (WRIG). Unlike the type I, power electronics are used to control the WRIG hence allowing a speed variation of up to 10%. As is the case with the Type I, Type II also has reactive power compensating components due to its inductive part [7].

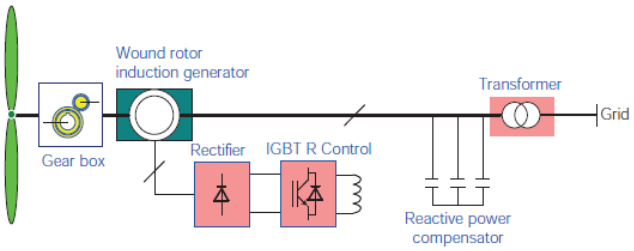


Fig. 5. Type II Wind Turbine Generator [1]

Type III: Doubly Fed Induction Generator
It is the most common technology used currently. It uses the WRIG that is fed to the grid through two paths. 40% of the power output passes through the converter to the grid while the remainder is directly fed to the grid. This gives the DFIG a flexibility of around 40% in speed variation [1]. Other than that, the power electronics technology used is more superior than in previous wind energy generators, it has decoupled active and reactive power control and hence better performance.

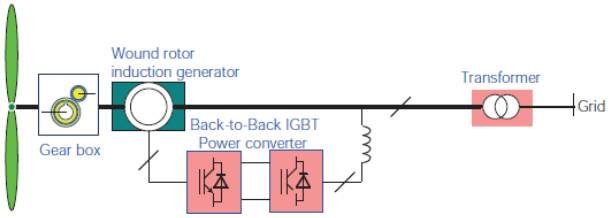


Fig. 6. Type III Wind Turbine Generator [1]

Type IV: Full Power Conversion Wind Turbine Generator
It is based on the SCIG and all its power is fed to the grid through the converter. It is this decoupling that allows it to have better control capability and little impact to the grid in the event of short circuit currents. It has similar properties to the Type 3 but achieves better speed and voltage control.

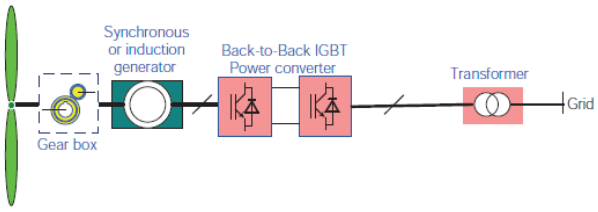


Fig.7. A Type IV Wind Turbine Generator [1]

C. Escalating Energy Access and Demand

There is a growing global demand for energy as people seek to get connected especially in developing countries. From 2009 to 2035, developing countries have shown and are projected to have a steady growth in demand of about 3.5% [1]. This demand for connectivity and hence load growth calls for injection of more power into the grid. The world is also looking at diversification of power generation portfolio with the aim of reducing reliance on fossil based power sources whose prices are not reliable and are prone to depletion. The world is going *green* with the aim of reducing its carbon footprint. Coal, oil and natural gas contribute the biggest share of CO_2 emissions amongst all generation technologies [8].

In 2009, the global energy consumption stood at 3900TWh. By 2035, it is projected that the figure shall be around 11100TWh which signifies a growth of around 300%. Fig. 7 below shows the growth of renewables from 2015 to 2035 for all renewable energy technologies [9]:

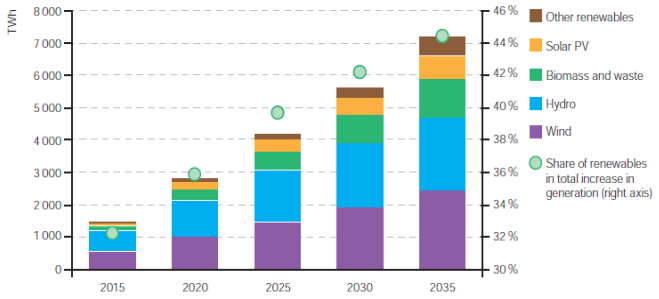


Fig. 8. Incremental Renewable Energy Based Generation Relative to 2009 [9].

The growing contribution of wind energy to the renewable energy mix from 2015 to 2035 can be noted in Fig.8 above. In 2015, wind was projected to contribute to 26% of the power generated through renewables. By 2035, it shall be contributing 35% of the renewable energy mix in the world [1]. This shows a very steady solid growth and adoption of wind energy technologies in a span of two decades hence showing the importance of this field to researchers. Power system stability and security is about to change as we know it and this transformation can be attributed to integration of renewables with stochastic behavior. Bulk wind power integration has taken the lead in having the highest capacity of installed renewables with intermittent behavior and hence forms the backbone of the review work in this paper.

Many countries have been very ambitious in the adoption of large capacity wind plants as shown in Fig. 8 below:

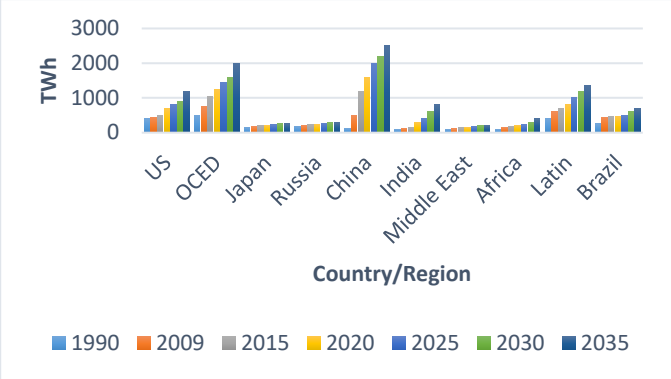


Fig. 9. Global Renewable Energy per Country/Region.

China intends to be the world leader in wind energy integration into the grid by 2035 as seen from their ambitious targets. In 2010 alone, China installed 16GW of wind energy while Europe (as a region) installed only 10GW. The bulk of the wind capacity in Europe was installed by Germany and Spain [10].

In 2009, 159GW of the global installed Wind capacity generated 273TWh of energy. It is projected that by 2020, the world shall generate 1282TWh from wind alone – 8 times what was generated in 2009. By 2035, the power generated from wind hits 2703 – double the 2020 projection. Capacity grows from 159GW in 2009, to 582GW in 2020, to 1102GW in 2035. A keen look at the statistics from 2009 to 2035 shows that compared to 2009, more power shall be generated for every kW of installed capacity. This is an indicator of wind technology and harnessing maturing through time into a more efficient form [1]. Fig. 10, Fig. 11 and Table I below summarize these observations:

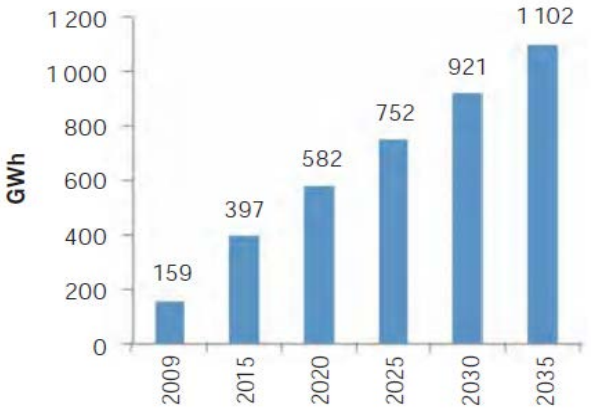


Fig. 10. Projected Global Installed Capacity for Wind Energy [9]

Fig. 11. Projected Global Wind Generation from Fig. 9 Installed Capacities

	2009	**2015**	**2020**	**2025**	**2030**	**2035**
Installed Capacity (GW)	159	397	582	752	921	1102
Generated Energy (TWh)	273	835	1282	1724	2182	2703
Generated Energy per GW of Installed Capacity (TWh/GW)	1.717	2.103	2.203	2.293	2.369	2.453

Table I: Generated Energy per GW of Installed Capacity

III. LARGE SCALE WIND INTEGRATION CHALLENGES

Wind is a naturally occurring resource that is intermittent in nature. Below are some of the challenges that system operators face when it comes to integration of wind power into the grid [11]:

Non-controllability – the fact that wind occurs naturally in nature means that it cannot be controlled by generation operators. Wind can only be consumed when available and when it is not available then operators have to look for other sources to satisfy the existing demand. The output of coal plants can be controlled by varying the quantity of coal used, hydros can be controlled by closing/opening of the valves which in turn controls the amount of water acting as the prime mover. This non-controllability nature of wind has in previous years not been much of a concern due to the limited injection of wind power into the grid. However, with the projected increase in the injection of bulk wind power into the grid this becomes a critical concern for power system operators. The balance of demand and supply becomes more sensitive since one of the large source in the grid does not inject constant power.

Partial predictability – Nature cannot be predicted with 100% precision and there is always a variation between the projected and the actual values. The output of wind cannot be determined to exact quantities. Despite all this, researchers have continued to invent and re-invent better systems aimed at prediction of the output of wind plants hence allowing the plants some form or partial dispatchability. This partial predictability if not well checked lead to voltage and frequency issues in the grid which affect the normal operations.

IV. THE FUTURE

It is now a global fact that power generation needs to move from being fossil-based to the renewable sources. Renewable energy is free in nature and despite its unpredictable nature can still be harnessed for the benefit of all. With fossil resources globally projected to get depleted within the decade, it is important to plan for the growing need for electrical energy by diversification of power sources for our grids. Wind energy is one of the sources that has the greatest potential for injection into the grid. The future is big; the future entails bulk wind injection into the grids and without proper planning, the world shall not be in a position to optimize harnessing of this free resource. Below is the future of the grid factoring in bulk wind injection:

Focus on generation flexibility: Conventional sources of energy have some form of generation flexibility in that the generation operator can vary the output of the generator when need be. The interaction of conventional generators with the grid falls under the control of utility operators hence output can be adjusted based on prevailing demand [12]. Wind, on the other hand, can only be consumed when available. The future lies in developing of control systems for wind farms and grids to ensure seamless integration of sources of power with various portfolios. Development of more efficient wind power prediction tools would also serve to give power system operators more specific information to allow for better planning and scheduling of the available resources [9].

Clustering of wind power plants: A good working example of such a setup is the Jiuquan Wind Power Base in China shown in Fig. 12 below [13]:

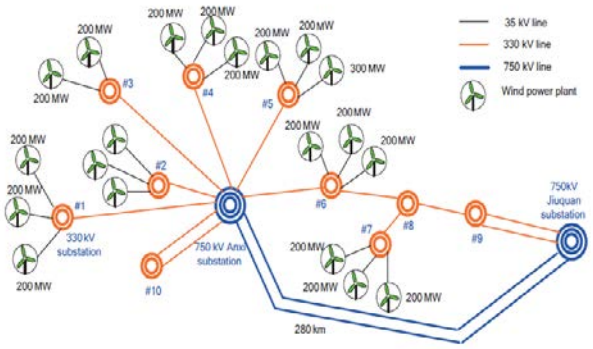


Fig. 12. Schematic of the Jiuquan Wind Power Base in China [11].

This scheme borrows from the concept of cascaded hydros on a river. The active and reactive power balances of the wind cluster scheme can be intelligently coordinated such that the output injected into the grid does not cause any problems to the wider grid. For example, in the event that one cluster is injecting reactive power and another is absorbing reactive power, the cluster arrangement ensures that the reactive power balances out within the scheme and the net reactive power injected into/absorbed from the grid is zero. This would in turn ensure better voltage control in the grid. Using a wide geographical location helps in smoothening out oscillations in the system compared to concentrating wind farms in one location. It is reported that wind farms located in different geographical locations are 33 – 47% more reliable compared to those clustered in one geographical location [12].

Pricing incentives for bulk consumers: Bulk consumers can be encouraged to consume in instances when wind power injected into the grid is at its highest. This would ensure that this free power is mopped from the grid and does not go to waste since the supply exceeds the demand. Active and reactive power balances would become easier with the concept of load shifting using price based incentives. Electric vehicles which are the future of automobiles can also be used as storage facilities which can mop up the excess wind energy say at night when the load curve experiences a dip and the wind energy is at its highest.

Interconnection of grids and transmission expansion: Transmission expasion and grid interconnection go hand in hand. Interconnection of two grids from two different countries would have a large wind power plant constructed and this resource shared. This would reduce the costs incurred to have units commited as spinnign reserve since this can be catered for by the generators in the two or more grids tied together.

V. Conclusion

Bulk wind power injection into the grid is considered one of the most favorable options in grids globally as the world seeks to shift to cleaner sources of energy and at the same time meet the growing energy needs. However, high penetration levels of wind energy into the grid shall introduce challenges in the operations of power systems. The stochastic nature of wind introduces another factor of uncertainty in power systems operation thus putting the integrity of the grid at stake. However, researchers and power system operators are continually developing strategies to mitigate these effects of variability of wind power to ensure that grid security is ascertained. With improved wind plants control technology, better forecasting techniques, grid interconnectivity and pricing incentives for consumption of wind power, the future can only be bright as the world taps into this free resource by nature.

VI. References

[1] IEC, "Grid Integration of Large Capacity Renewable Energy Sources and use of Large Capacity Electrical Energy Storage," International Electrotechnical Commission, Geneva, 2012.

[2] Z. Chen, "Issues of Connecting Wind Farms into Power Systems," in *IEEE/PES Transmission and Distribution Conference & Exhibition*, China, 2005.

[3] Y K. Wu, L. T. Chang, S. M. Chang, T. Y Hsieh and T. C. Lee, "Integration of Large Scale Renewable Power into the Taiwan Power System," in *IET Conference*, 2015.

[4] Shahab Shokrzadeh, Mohammad Jafari Jozani, and Eric Bibeau, "Wind Turbine Power Curve Modeling Using Advanced Parametric and Nonparametric Methods," in *IEEE Transactions on Sustainable Energy*, 2014.

[5] Anissia Beainy and Chantal Maatouk, "Comparison of Different Types of Generator for Wind Energy Conversion Systems Topologies," in *International Conference on Renewable Energies for Developing Countries (REDEC)*, 2016.

[6] Jason M. Kennedy, Brendan Fox, Tim Littler and Damian Flynn, "Validation of Fixed Speed Induction Generator Models for Inertial Response Using Wind Farm Measurements," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1454 - 1461, 2011.

[7] Eduard Muljadi, Mohit Singh, and Vahan Gevorgian, "Fixed-Speed and Variable-Slip Wind Turbines Providing Spinning Reserves to the Grid," 2013.

[8] World Nuclear Association, "Comparison of Lifecycle Greenhouse Gas Emissions of Various Electricity Generation Sources," London.

[9] International Energy Agency, "World Energy Outlook 2011," 2011.

[10] IRENA-GWEC, "30 Years of Policies for Wind Energy: Lessons from China," pp. 48 - 52, 2012.

[11] L.Bouzidi, N. Bellaaj and M. El Euch, "Large Scale Wind Power Integration : extended power system operations," 2015.

[12] T.R. Ayodele, A.A. Jimoh, and J.T Agee, "Challenges of Grid Integration of Wind Power on Power System Grid Integrity: A Review," *International Journal of Renewable Energy Research*, vol. 2, no. 4, pp. 618 - 626, 2012.

[13] Hong Shen, "Transmission of large scale wind power in China," *IEC Workshop on Grid Integration of Large-capacity Renewable Energy and Large-capacity Energy Storage*, 2012.

The evaluation of the impact of Inter-Cell Interference Coordination on the performance of users in an LTE system

Daberechi Goodnews Uwakwe, Ayodeji Akeem Ajani, Odum Rowani, and Anthonia Nwakaego Isaac-Ugbogu

***Abstract*—**The Long-Term Evolution (LTE) mobile/wireless standard was introduced with the motivation that it would offer remarkable improvement to the previous communication standard – Evolved High Speed Packet Access (HSPA+). LTE systems, unlike the earlier standards, tend to utilize the available frequency spectrum in each cell of the network and hence promise to offer higher throughput to the users in the network, better system capacity, lower latency and delay, improved spectral efficiency etc.

In order for the standard to effectively meet up with these performance targets, it has to eliminate or minimize the interference on the network. LTE adopts the Orthogonal Frequency Division Multiple Access (OFDMA) method which successfully eliminates the presence of intra-cell interference by enabling the users in each cell to transmit orthogonally. However, the standard still suffers inter-cell interference which could be as a result of two cell-edge users located in two adjacent cells communicating at the same frequency or both causing interference to each other due to the high-power level at which they transmit. This leads to an overall reduction in the system performance in terms of signal to interference and noise ratio (SINR) values, system capacity, users' potential data rates.

This research studies the performance of cell-centre and cell-edge users in a 7-cell LTE cellular network model as simulated on MATLAB 7.11. This research evaluates the performance in terms of SINR, capacity and spectral efficiency of the users in the network with more focus on the users in the cell-edge region. A comparison was made of the performance of the users located in the reference cell in two conditions; when the interference from all the adjacent cells was not managed; and when inter-cell interference coordination (ICIC) was implemented. ICIC implementation was by shutting down one or more interfering cells. The comparison was to show the level of improvement as perceived by the users in the cell with the implementation of ICIC.

The outcome of the study showed improvement in the cell users' quality of experience (QoE) which includes higher SINR values, increased capacity and better spectral efficiency in the network as the number of sources of inter-cell interference reduces.

***Keywords*—**Interference, Inter-Cell Interference Coordination (ICIC), Long-Term Evolution (LTE), MATLAB, Orthogonal Frequency Division Multiple Access (OFDMA), Signal to Noise Ratio(SINR).

I. Introduction

As the necessity for mobile broadband increases, improvement on existing mobile communications standards is imperative to provide the data and voice services required by mobile and wireless devices and the evolving data hungry applications.

Interference being the major challenge in LTE systems, has led to significant number of research works in this area of study. Inter-cell Interference (ICI) occurs when users in different neighbouring cells make attempts to use the same radio resource(s) at the same time or users in the neighbouring cells communicate at high power such that their signals act as interference on the users in the other cells [1]. Fig. 1 shows a graphical demonstration of ICI whereby a UE which moves further away from its serving eNodeB A experiences inter-cell interference from an adjacent eNodeB B which possibly allocates the same frequency channel f1 to UEs in its cell. This happens because each of the cells only knows what radio resources its UEs use and hence they independently schedule and allocate the same frequency resource to their UEs.

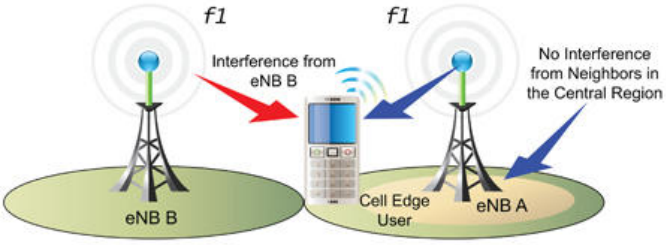


Fig. 1 A demonstration of Inter-cell interference in an LTE system [2]

D. G. Uwakwe, School of Computing Science and Engineering, University of Salford, UK (e-mail: daberechiuwakwe@gmail.com).
A. A. Ajani, Department of Electrical Engineering, PAUISTI (phone: +254743126864; e-mail: ajaniaa@gmail.com).
O. Rowani, School of Computing Science and Engineering, University of Salford, UK (e-mail: R.Odum@edu.salford.ac.uk).
A. N. Isaac-Ugbogu School of Computing Science and Engineering, University of Salford, UK (e-mail: toniadear@gmail.com).

II. LITERATURE REVIEW

A. *Interference in LTE systems*

While LTE and later standards succeed in achieving their intended targets or specifications [3] by adopting tools such as the frequency reuse [4] of 1 (that is the use of the same frequency channel in all the cells in the network) and also creating multilayer heterogeneous networks (in which smaller cells are created within the coverage area of macro cells), they also bring about one of the most significant challenges in the mobile and wireless communications industry today– Interference, by increasing the portion of cell-edge users where two or more base stations compete for coverage and can transmit to and receive from the same user equipment (UE) device [2].

The effect of this interference results in a lower Signal to Interference and Noise Ratio (SINR), a degradation of network performance and user experience, and a diminished efficiency of use of network resources. Some of the interference however can be prevented by some careful Radio Frequency (RF) planning, but interference in networks cannot be completely eliminated [5].

In LTE networks, two different kinds of interference could be experienced by users and these are the intra-cell interference and inter-cell interference [6]. These two kinds of interference contribute to the degradation in the SINR experienced by the users in the cellular network. This can be seen in the equation shown below which is derived from a cellular network based on a Fractional Reuse Factor of n (FRF-n).

$$SINR, \quad Y^n = \frac{P_{desired}}{P_{intra-cell} + P_{inter-cell} + \frac{P_{noise}}{n}} \qquad (1)$$

Where:
Y^n: Signal to Interference and Noise ratio (SINR)
P_desired: Power of the desired user's signal
P_(intra-cell): Power of the intra-cell interference
P_(inter-cell): Power of the inter-cell interference
P_noise: White noise power
n: Frequency Reuse Factor (FRF)

From equation (1), it can be seen that the SINR experienced by users in a cellular network is limited by the intra-cell interference, inter-cell interference and white noise. [7]

B. *Inter-cell Interference (ICI)*

In a cellular network layout, the cells are tessellated across the network with each cell having a transceiver which consists of either omni-directional or directional antennas covering each cell. This concept of splitting the network into cells brings about the idea of categorising users into cell-edge users (CEUs) and cell-centre users (CCUs) [8]. The CCUs are located well within the coverage of the cell's transceiver while on the other hand; the CEUs are located at a point where two or more cells on the same cellular network overlap or very close to each other.

When two cells overlap, a UE at the cell-edge could be receiving signals from two or more of these contiguous cells and thereby resulting in Inter-cell interference [8]. This kind of interference could also be experienced in a heterogeneous network whereby UEs receive signal from both the macro and small cells within the network [9]. Fig. 2 is an illustration of inter-cell interference when two cells overlap and when it occurs in a heterogeneous network.

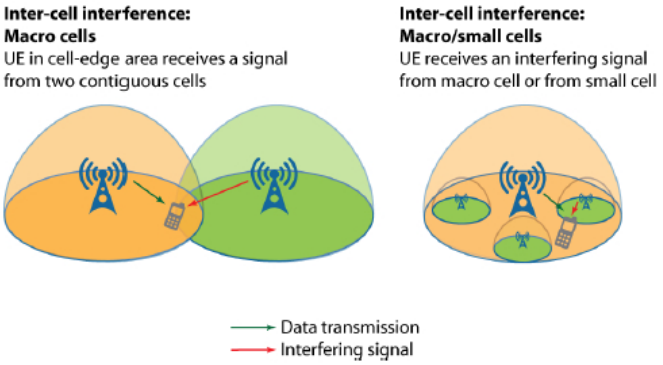


Fig. 2 Inter-cell interference between users in different but adjacent cells. [2]

Since there are limited spectrum resources, most mobile operators deploy in their LTE networks a frequency reuse = 1 configuration. This deployment scheme is known as a Single-Frequency Network (SFN) and it simply means that a single carrier frequency is reused in all cells of the network [10]. SFNs are commonly used in LTE networks to effectively utilize the limited radio spectrum and to also increase the coverage area. However, SFNs by nature are limited by inter-cell interference [10].

III. SPECIFICATION AND DESIGN (METHODOLOGY)

A. *Requirement Phase*

In order to show the impact of inter-cell interference from interfering cells on the users in a particular cell, a simulation model that consists of a seven hexagonal cell sites is considered. Each cell has an eNodeB assumed to be located at the centre of the cell; and each of these eNodeBs is equipped with an omni-directional antenna all transmitting at the same power level. The hexagonal cell layout has been adopted due to its conceptual and computational simplicity to approximate the cells which are irregular and complex in real life as a result of terrain features and artificial structures. The cellular network model is as shown in Fig 3.

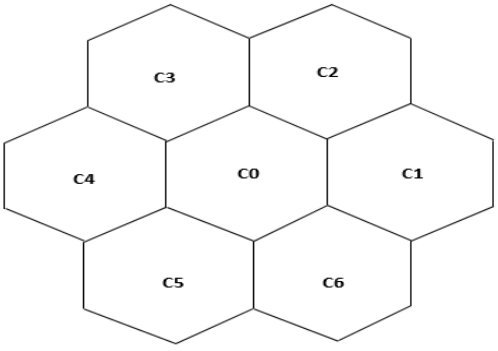


Fig. 3 Hexagonal Cellular Network Layout

The inter cell site distance has been chosen to be 100m and the UEs are randomly located between the eNodeB and the edge of the cell. The cell centre UEs are randomly located at a distance of about 10 – 15m from the eNodeB whereas the cell edge UEs are located at a randomly generated distance of about 80 – 85m from the eNodeB. A reference distance of 10m has been chosen and this forms a basis on which the path-loss gain for each UE is computed.

In this project, a free space signal propagation (which has a path-loss exponent of 2) has been assumed in which the signals tend to travel from the transmitter to the receivers without encountering any obstacles. However the users experience a distance dependent path-loss which impacts on the implementation of this project. The path-loss is defined as the loss in signal strength from the transmitter to the receivers as a result of a line-of-sight (LOS) path through free space. The path-loss gain has been employed in the computation of the received power for each of the receivers by subtracting its value from that of the transmitter power of the eNodeB.

The antennas of both the eNodeB and the UEs are both assumed to be omni-directional and since there is no potential danger of multipath effects in a free space propagation environment, the Single input, Single out, SISO antenna configuration has also been adopted. SISO refers to a wireless communications system in which one antenna is used at the source (transmitter) and one antenna is used at the destination (receiver).

For convenience, the system and simulation parameters used in this project are summarized in Table I;

TABLE I
SYSTEM AND SIMULATION PARAMETERS

PARAMETER	Value
Cellular Layout	Hexagonal
Cell Radius	100m
Path-loss exponent	2 (Free space propagation)
Antenna Configuration	SISO
Reference Distance d0	10m
eNodeB transmit power	1dBi
Noise figure	5e-16
Bandwidth of a RB	1e3
Cellular Layout	Hexagonal
Cell Radius	100m

B. *Chosen Approach and Simulation Scenarios*

The approach of this study shows how significantly the inter-cell interference in a cellular network affects the cell-edge and cell-centre located users in a particular cell. The approach also shows how avoiding or mitigating the inter-cell interference could impact on the performance of the users.

A cellular network may in real life consist of hundreds or even thousands of cells in a particular region of LTE deployment. However, this research considers only a few cells to keep it simple and manageable.

In order to show a good level of variation in the simulation results, different designs and scenarios are developed each of which depicts one of the following;

a) The users' experience when there is only one cell and hence no source of interference in the network.

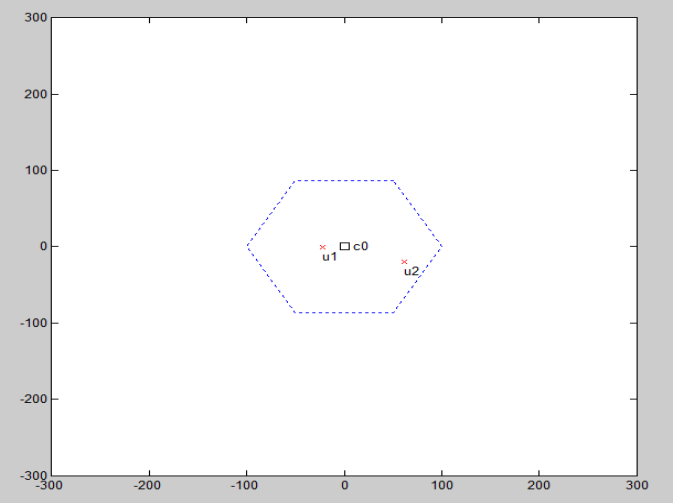


Fig. 4 Two (2) users in a single cell scenario

b) The users' experience when 5 UEs are randomly located in a single cell scenario

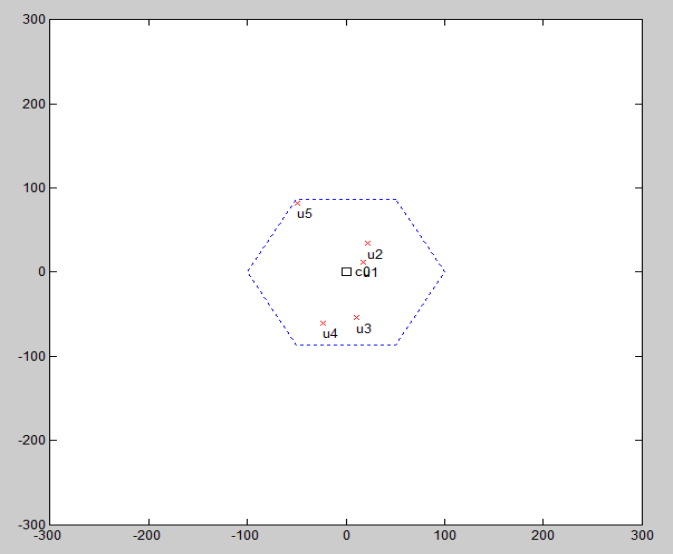


Fig. 5 Five (5) users in a single cell scenario

c) The degradation of the users' performance in (a) when two sources of interference are added to the network. This actually is close to a real-life case than (a) since no cell exists in isolation in practice.

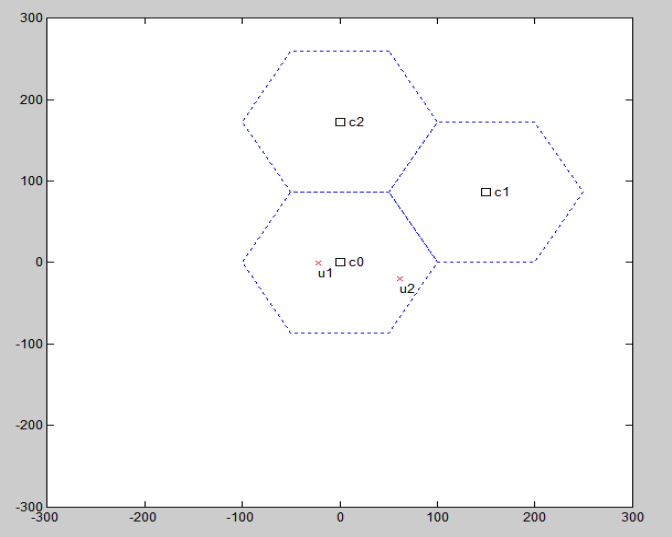


Fig. 6 Two (2) users in a multi-cell (3 cells) scenario

d) More sources of inter-cell interference (ie cells) are added to (b) so as to see how much more degradation in performance the users experience.

e) Finally, more UEs are added to our cell of interest at randomly generated positions between the eNodeB and the edge of the cell. The sources of interference are also increased to 6 cells. The output of this shows a very clear picture of users' performance based on their location in the cell.

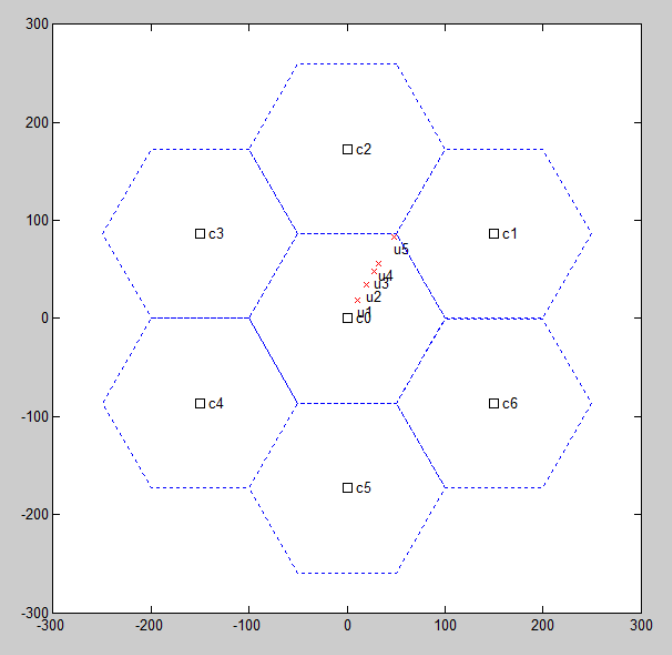


Fig. 7 Five (5) users in one cell of a multi-cell (7 cells) cellular network

IV. DEVELOPMENT AND IMPLEMENTATION

A. *MATLAB algorithm development*

The MATLAB algorithms developed in this study have been built in a modular fashion, with each module being responsible for the execution of a discrete aspect of the desired system functionality which is to study the performance of user equipment (UEs) in the cell-centre and cell edge regions. In this work, the following have been designed and implemented as a section of the algorithm but together they add to the overall success of this study.

a) Distance of the UEs from the eNodeB

In this study, the metrics measurements taken have been done only in the reference cell which is the middle cell in the cluster as defined in the model. The users whose performance metrics are measured have been designed to be located at a randomly generated distance between the eNodeB and the edge of the cell. For the first scenario, the distances of U1 (cell-centre user) and U2 (cell-edge user) from the eNodeB are calculated with the following section of the algorithm;

```
Ranges = sort (d0+(R-d0)*rand(U,1), 'ascend');
```

In this line of code, rand(U,1) generates the random variable which helps generate random positions for the UEs U1 and U2.

For the scenario with more than just the two UEs, the users' positions are also randomly generated using the line of section of the code;

```
Ranges = sort (d0+(R-d0)*rand(U,1), 'ascend');
```

The MATLAB function "sort" and "ascend" have been used to ensure that the smallest random distance is assigned to U1 which has been assumed to be closest to the eNodeB and the largest random distance assigned to U5 which is the supposed farthest UE from the eNodeB.

b) Angular position of UEs in the cell

While the UEs used in this simulation are distributed at random distances between the eNodeB and the edge of the cell, these UEs are also located at an angle from the eNodeB. Their angular positions however do not affect the signal transmission/reception since the antennas are omni-directional. The users' angular positions have been randomly generated with the following lines of codes;

```
Angles = 2*pi*rand(U,1); For the scenario with just 2 users

Angles = pi/3; For the scenario with more than 2 users
```

Combining the UEs' distances from the eNodeB and their angular position, their actual positions in the cell have been defined with the following line of code;

```
uPos = [Ranges.*cos(Angles),Ranges.*sin(Angles)];
```

c) Path-loss gain of the UEs

The path loss includes all the lossy effects associated with the signal propagation distance between a transmitter and a receiver. It is basically the reduction in power density of an

electromagnetic wave from a line-of-sight (LOS) path as it propagates through free space. It does not include any loss associated with hardware imperfections and it assumes that the antenna gain is 1.0 or 0 dBi. In this work, the path loss gain for UEs in the cell has been computed using the following line of MATLAB code;

```
PL = @(d) 128.1+10.*n.*log10(d/d0);
```

Where
n = path-loss exponent for free space propagation.
d = the distance of the UE from the transmitting antenna
d0 = the reference distance
128.1 = a constant known as the path-loss constant

d) UEs' receive power

The receive power is the signal strength of the transmitted electromagnetic waves as measured at the receiver's antenna. It is equal to the transmitting antenna's output power minus the path loss (attenuation) as the signal propagates from the transmitter to the receiver. In short, RX input power (dBm) = TX output power – path-loss gain (field attenuation). This is executed for each UE's antenna using the MATLAB line of code;

```
uRxPow = P0.*10.^(-0.1*uPL);
```

e) SINR for each UE

The Signal to Interference and Noise Ratio (SINR) is a quantity used to give theoretical upper bounds on the rate of information transfer (channel capacity) in wireless communications networks. SINR is simply defined as the power of a certain signal of interest divided by the sum of the interference power from all other interfering signals and the power of the background/thermal noise. For the UEs, their SINR has been computed using the MATLAB code;

```
uSINR(:,i+1) = uRxPow(:,1) ./ (N0B + IPow);
```

The numerator on the RHS of the equality sign computes the power of the signal of interest (signal from the central cell's eNodeB); the denominator on the other hand sums up the power of the signals from the interfering cells and the system noise.

f) Channel Capacity for each UE

The channel capacity is the tightest upper bound on the rate of information that can be reliably transmitted over a wireless communications channel. The channel capacity is given in bits per second and has been calculated in this research in accordance with the Shannon-Hartley theorem using the MATLAB code;

```
uCap = B*log2(1+uSINR);
```

uSINR is the signal to interference and noise ratio as derived for each user above and B is the bandwidth of the channel in Hertz.

g) Spectral efficiency for each UE

Spectral efficiency refers to the information rate that can be transmitted over a given bandwidth in a specific communication system. The spectral efficiency for each of the UEs is a function of their SINR and it has been implemented on MATLAB as follows;

```
uSpec = log2(1+uSINR);
```

The MATLAB algorithm showing how all the above components and other bits were merged to implement this design in this project is as shown in appendix A.

V. RESULT, DISCUSSION AND ANALYSIS

The users' performance has been measured using the metrics Signal to Interference (and Noise) Ratio (SNR/SINR), Capacity and Spectral Efficiency. These measurements have been taken only for users in the central cell which is this research point of focus.

A. The simulation scenarios and the results

The scenarios chosen to implement this project are presented here starting from the simplest one which is of a single cell and a 2 UEs. The size of the scenarios is increased in each step so as to present a view of the network performance when more users and cells are added to the network.

These scenarios, their results and discussions are as follows;

a) Single cell with 2 UEs at different positions

In this scenario, a single cell was modelled which has only two UEs, one located within the cell-centre distance range of the cell and the other located at a cell-edge distance range of the cell. Being the only cell in the network, it is assumed that there is no inter-cell interference coming from other cells and therefore the two UEs user experience are measured based on their position only. The results obtained in this scenario are as follows;

TABLE II
PERFORMANCE OF 2 USERS IN A SINGLE CELL CELLULAR NETWORK

	SNR (dB)	Capacity (b/s)	Spectral Efficiency (b/s/Hz)
UE1 (cell-centre)	66.28225	6072.154	4.208897
UE2 (cell-edge)	7.460729	3080.782	2.135435

From Table II, the two users UE1 and UE2 based on their distance from the eNodeB and hence path-loss gains, have different SNR, capacity and spectral efficiency. This shows that the distance of users from the centre of the cell is a determinant of their performance.

b) Single Cell with 5 UEs at different positions

In this scenario just like in the previous one, only one cell exists in the network but this time the number of users have been increased to 5 and each are place at random distances from the eNodeB. On implementation of this, the following results in Table III were obtained

TABLE III
PERFORMANCE OF 5 USERS IN A SINGLE CELL CELLULAR NETWORK

	SNR (dB)	Capacity (b/s)	Spectral Efficiency (b/s/Hz)
UE1(closest to eNodeB)	66.28225	6072.154	4.208897
UE2	19.06931	4326.919	2.999192
UE3	10.10888	3473.642	2.407745
UE4	7.460729	3080.782	2.135435
UE5 (closest to cell-edge)	3.393509	2135.374	1.480128

As shown in Table III, the UEs experience some level of degradation as their distances from the eNodeB (cell-centre) increases. This is clearly shown in the performances of UE1 and UE5 in which UE1 being the closest to the eNodeB has SINR, capacity and spectral efficiency values of approximately 66dB, 6072b/s and 4b/s/Hz respectively whereas UE5 which is the farthest from the eNodeB has SINR, capacity and spectral efficiency values of approximately 3dB, 2135b/s and 1b/s/Hz respectively.

c) Single cell with 2 UEs + 2 interfering cells

In this scenario, two cells which would be sources of inter-cell interference are added to the cellular network in scenario i. The two users are as usual placed randomly in the first cell with UE1 at the cell-centre region and UE2 at the cell edge region. On implementation, the result showing the performance of the UEs with respect to the number of interferers is as shown as follows in Table IV.

TABLE IV
PERFORMANCE OF 2 USERS IN A MULTI CELL (3 CELLS) CELLULAR NETWORK

	No of Interferer(s)	SNR (dB) (b/s)	Capacity	Spectral Efficiency (b/s/Hz)
UE1	0	66.28225	6072.154	4.208897
UE2	0	7.460729	3080.782	2.135435
UE1	1	36.09733	5213.244	3.613545
UE2	1	2.851968	1945.596	1.348584
UE1	2	23.28352	4601.906	3.189798
UE2	2	2.212733	1683.801	1.167122

d) Single cell with 5 UEs in cell C0 + 6 interfering cells

In this scenario, 4 more sources of inter-ell interference are introduced by adding 4 cells adjacent to cell c0. 3 more UEs are also placed in cell c0, making it a total of 5 UEs and all are placed at random distances from the eNodeB with UE1 being closest to the eNodeB and UE5 the farthest. The UEs are located on the same plane at an angle of pi/3 (180o/3) from the eNodeB, this has been done so that all the UEs are given a fair ground for the evaluation of their performances.

Graphically, the SINR of the 5 UEs in the scenario above changes with the distance of the UE from the eNodeB and the number of sources of inter-cell interference as shown in Fig. 8 below.

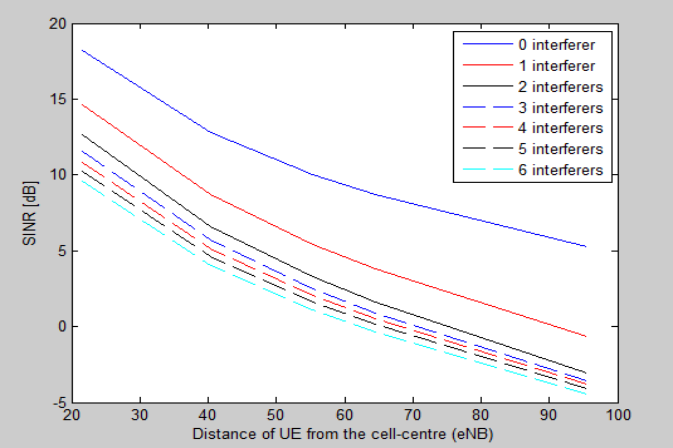


Fig. 8 The UEs' SINR values as influenced by their distance from the eNodeB and the number of interferers in the network.

From the Fig. 8, it can be clearly seen that the SINR values of the UEs suffers degradation as the number of sources of inter-cell interference increases. The long tail of the SINR distribution tends to the negative for users at a significantly long distance form the eNodeB and the tails tend even more to the negative when there are interferers in the network. This is noticed from the fact that the UEs have better SINR when there is no interferers in the network and then the values start dropping when the first interferer in introduced and continues almost at the same rate for each new source of interference added to the network. From the figure, it is also clear that the UE closest to the cell-centre (eNodeB) has better SINR values each time the number of interferers is increased than the UEs farther away from the eNodeB.

Fig. 9 is a graphical representation of how the spectral efficiency values of the UEs change with any change in their distance from the eNodeB and the number of sources of interferers.

The spectral efficiency which is in direct proportion to the SINR tends to exhibit the same behaviour as the SINR values of the UEs. As the distance between the UEs and the eNodeB increases, the spectral efficiency of the UEs drops at almost the same rate with the worst value observed in the UE farthest away from the eNodeB or closest to the cell edge. Similarly, the spectral efficiency of the users is best for each user when there is 0 interferer (as represented with a thick blue curve in the Fig.

9) in the network, but these values keep decreasing for each UE as the number of sources of interference increase.

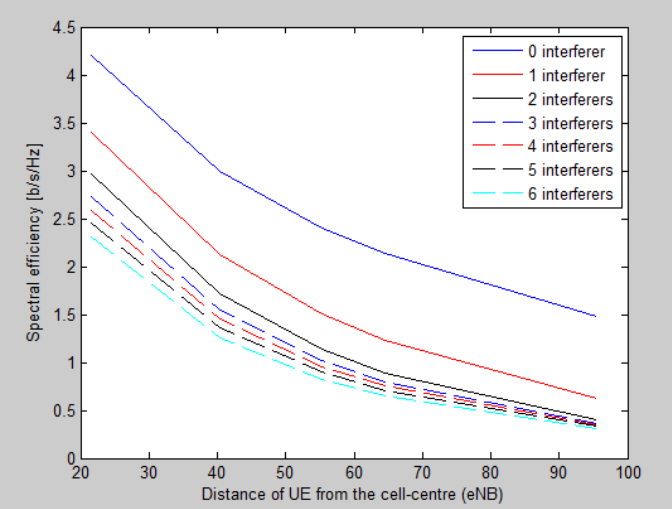


Fig. 9 The change in the spectral efficiency values of the UEs as their distance from the eNodeB changes and the number of interferers increase.

The results in this simulation have been obtained from the users in the reference cell – which is surrounded by other cells in a 1-cell, 3-cell and 7-cell cellular networks as implemented in this work. The results were obtained from an LTE system which uses a frequency reuse of N =1 whereby every cell is an interferer. The results explain the "Pilot Pollution" or "no dominant server" problem which describes a situation where power transmitted from many different cells appears in a location and none happens to be significantly better than the others. The UEs treat the best received power from one cell as "signal" and then power from other cells are treated as interference. For the cell-edge UEs therefore the composite signal is high but there is a very poor SINR from any single cell because the amount of the total signal treated as interference is too high.

B. Impact of shadowing effect on the performance of the users

In the course of propagation, a radio wave does not only attenuate through distance as is noticed in the case of path loss, but it does attenuate also through some physical phenomena such as scattering, reflection etc. depending on the type of environment observed. In this work, assumption has been made of a free space propagation which implies that there is a clear line of sight with no obstacles between the transmitter and the receivers. Aside the path loss attenuation, another propagation loss which could impact this project is shadowing – which is an effect of obstruction in the wave propagation. If shadowing had been considered in this work, the effect would lead to the fluctuation of the SINR values of the UEs as depicted in the Fig 10 below.

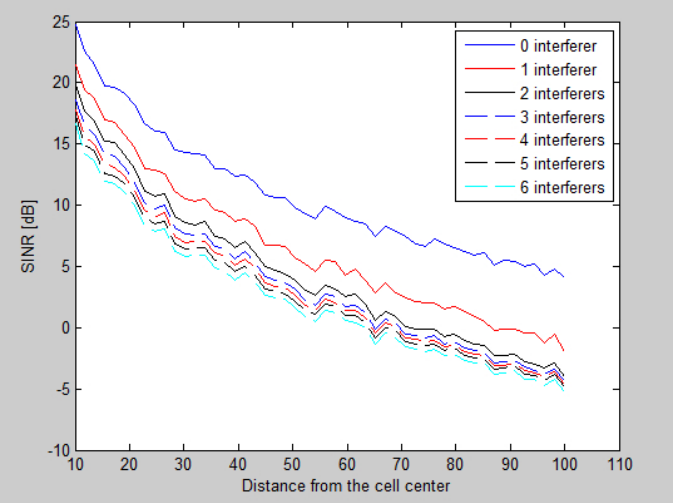


Fig. 10 Shadowing effect on the performance of the users with respect to their distances

C. LTE system performance in this model as perceived by the end-users and the system operators.

The network operators and the users both see the network differently depending on what roles/services each stands to play/receive. The system operators seek to provide efficient service to the users while at the same time keeping the system running. The users on the other hand look forward to receiving a certain level of quality of experience from the network.

a) The potential network capacity of this model

The network model in this project considered a maximum of 5 users randomly placed in the cell but in a real-life network, there is bound to be much more users in one cell at a time. Therefore, for this reason, attempt is made in this section of this work to find out how much UEs can exist in the cell as observed in this model.

In an attempt to achieve this, an assumption is made of the threshold SINR of 7dB with which the cell-edge UEs are distinguished from the cell-centre UEs. Assumption is also made of a threshold SINR value of 1dB below which the user is taken to have a zero performance in the network. When more UEs, precisely 100 are randomly added to the cell in the cellular network, it is noticed that about 25 of the UEs fall in the cell-centre region and the rest fall in the cell-edge region. And out of the 75 UEs in the cell-edge region, only about 45 show a good level of user performance by having SINR values greater than the SINR threshold which is 1dB. These measurements are taken from the reference cell when all the interferers are enabled. When the number of interferers is reduced, a drastic rise in the performance of the users is noticed. It should also be noted that there is also an increase in the number of users allocated to the cell-centre region as the number of interfering cells increase.

We can therefore conclude tentatively that given the assumption of the SINR values made in this model, the number of UEs likely to have a quality performance in the network when all the adjacent sources of interference are turned on would be within the range of 60-80 UEs.

b) The users' potential data rates

This is a major factor to be considered in a communications network particularly from the users' perspective. From Table III where 5 users are considered in a multi-cell network, it is noticed that the users' performance is better when the interference is coordinated (that is when at least one interferer is shut down). With more users on the network, there is bound to be better performance for the users classified as cell-centre users. The UE classification therefore plays an important role in the performance of the system since more users (those at the cell-centre) are able to have a good level of quality of experience.

The SINR threshold configured and which is a criterion for the classification of the users in the cell is very sensitive to the performance of the network. If the SINR threshold is reduced from the assumed 7dB in the previous section to say 5dB, more users would fall in the cell-centre region and the more cell-edge users would also have better throughput. When this SINR threshold is configured, the data rates of the users in this model fall within the range 2 – 8Mbps and 0.8 – 4Mbps for the cell-centre and cell-edge users. This depends however on the level of interference on the cell and the exact position of the UE. These are practical data rates observed for the LTE system in this scale of the model and it is as expected quite lower than what LTE theoretically offers because some factors such as interference, fading (path loss, shadowing), signaling, terrain etc has not been put into consideration when coming up with the theoretical values.

The fluctuations in the SINR values (performance in general) of the users are as a result of Pilot Pollution whereby there are adjacent cells and all are transmitting at about the same level of signal power. The UEs in the reference cell see all the signals all at once and each signal acting as interferer to each other.
To solve this problem, engineers tend to make attempts to eliminate the unwanted or interfering signals, by setting power parameters or physical adjustments (tilt, azimuth) so as to make only the dominant signals visible to the UEs.

D. CRITICAL EVALUATION.

In this research, the focus has been on the performance of cell-edge and cell-centre located users in an LTE system based on the impact of the inevitable inter-cell interference in the system. With the results obtained from the study, it has been shown how the SINR, spectral efficiency and potential capacity of the users at each location change with their distances from the eNodeB and the number of sources of inter-cell interference in the system.

This research has shown comparatively similar results when compared with previous research works in this area. Similar to the work by Islam and Chowdry (2013), the performance of the users in this work was observed to degrade with an increase in the amount of interference from neighbouring cells. There is even further performance degradation for users farther away from their serving eNodeB due to a further reduction in the SINR.

This research shows as results the values of SINR, spectral efficiency and capacity which show the relationship between the performances of the users to their position in the cell. Even with the above results, there is no apt explanation of the users' performance based on the metrics such as Bit Error Rate (BER), delay, jitter and latency which are important in the quantitative measurements of Quality of Service (QoS) in a cellular network. These metrics were not retrieved in this work as there was no initiation of traffic flow, protocol definition and/or channel allocation for the users as modelled in this project.

The values of the SINR, spectral efficiency and capacity of the UEs' tend to show us a trend observed in LTE systems for users randomly located in a cell that is affected by inter-cell interference. These results might not be perfect to be relied on because in a large scale deployment of the system, the experience of the cell-edge users could be worse off owing to the irregularities from the real-life cellular networks. These irregularities could also result in better values of SINR, spectral efficiency and capacity for users in some regions and poor values for users in some other regions within the same cell.
Lastly, though not considered in this research, the result of this study could also have been influenced heavily by the difference in the antenna heights of the transmitters (eNodeBs) and receivers (UEs).

VI. CONCLUSION

This research focuses on the evaluation of the performance of both the cell-edge and cell-centre users in an LTE network relative to the impact of inter-cell interference from the neighbouring cells. This research successfully implemented an LTE cellular network model in which the neighbouring cells in the defined cluster act as sources of inter-cell interference to the users in the reference (centre) cell of the model. The results also have shown how much impediment the interference from other users in other cells could cause on the performance of users in an LTE system.

In conclusion this research was able to answer the following questions;

- How much impact does inter-cell interference have on the users based on their locations in the cell of the LTE network?

The users as used in the model were observed to react negatively to the impact of the interference with a reduction in the values of the performance metrics as measured in the reference cell. The effect of the interference was felt more by the UEs located close to the edge of the cell as compared to the UEs in the cell-centre region whose performance were somewhat better and acceptable within the LTE requirements.

- Does an ICIC LTE network perform any better than an LTE system in which ICIC is not implemented?

As depicted by the results of the simulations, the isolated cell in which there was no effect of inter-cell interference showed a remarkable level of quality of experience in the measurements taken. This scenario which represents a case where ICIC is wholly implemented, successfully showed how much better the users perform when ICIC is implemented as compared to when it is not.

- What is the impact of the inter-cell interference on system capacity of the LTE system modelled in this work?

This work has also shown that in LTE systems, the system capacity which is defined as the number of users the system can

service, is inversely proportional to the level of inter-cell interference experienced in the system. This however depends on the SINR threshold value which determines what level of performance is acceptable in the system. With a choice of low SINR threshold, more users tend to show good performance in the network. The number of sources of inter-cell interference also impacts the system capacity as it was seen that with an increase in the number of interferers, the system capacity (number of UEs with acceptable performance) increases.

In this research, the strategies employed in inter-cell interference avoidance were also studied and the pros and cons of each technique were highlighted.

ACKNOWLEDGEMENT

The authors wish to appreciate the members of the telecommunication group at the University of Salford. We are also grateful to our friends and family for their support.

REFERENCES

[1] Michelle M. D., & Harrison J. S. (2014). Interference Coordination in LTE/LTE-A (1): Inter-Cell Interference Coordination (ICIC). Retrieved on 14th December 2014 from https://www.netmanias.com/en/post/blog/6391/lte-lte-a/interference-coordination-in-lte-lte-a-1-inter-cell-interference-coordination-icic

[2] Paolini, M. (2012). Interference Management in LTE networks and Devices. Senza Fili Consulting.

[3] 3GPP TR 25.814 retrieved from http://www.qtc.jp/3GPP/Specs/25814-710.pdf on 22nd January, 2018

[4] Xie Z, Walke B. Frequency reuse techniques for attaining both coverage high spectral efficiency in OFDMA cellular systems. In: IEEE WCNC; 2010. p. 1–6.

[5] LTE best practices", *CommScope*, March 2015.

[6] Esteves, H., Pereira, M., Correia, L.M. and Caseiro, C., 2006. Impact of intra-and inter-cell interference on UMTS-FDD. Graduation Thesis, Instituto Superior Técnico, Lisbon, Portugal.

[7] W. E. I. Xu, X. Wu, X. Dong, H. U. A. Zhang, and X. You, “Dual-Polarized Massive MIMO Systems Under Multi-Cell Pilot Contamination,” IEEE Access vol. 4, 2016.

[8] Lee, D., Li, G. Y., & Tang, S. (2012). Inter-Cell Interference Coordination for LTE Systems. IEEE Global Telecommunications Conference, pp 4828 – 4833.

[9] Lopez-Perez, D., Guvenc, I., De La Roche, G., Kountouris, M., Quek, T. Q., & Zhang, J. (2011). Enhanced intercell interference coordination challenges in heterogeneous networks. Wireless Communications, IEEE, 18(3), 22-30.

[10] Sequans Communications (2012). Mitigating Interference in LTE Networks with Sequans AIR™ - Active Interference Rejection

Estimation of Cropping Regimes using High Temporal Frequency Moderate Resolution SITS Analysis

Nduati E.[1], Jong Geol Park[2], Wei Yang[1], Akihiko Kondoh[1]
[1]Center for Environmental Remote Sensing (CEReS), Chiba University, Japan
[2]Tokyo University of Information Science, Japan

***Abstract*—**Horticultural crops are economically, environmentally and nutritionally beneficial and are more intensively cultivated than field crops. However, there are many varieties, they have shorter growing periods and are cultivated on small parcels of land. The type and variety of crops, while limited by an area's weather and soil conditions, is also at the discretion of the farmer and is influenced by market forces. This spatially and temporally dynamic nature means that annual crop type mapping does not adequately provide the information necessary for forecasting. To plan and allocate resources, a higher frequency of crop inventorying involving identification of what crops are grown, and where and when they are grown is necessary. Remote Sensing and Geographical Information Systems provide excellent tools for the collection and analysis of vast amounts of data. Further, high temporal frequency data lends itself to monitoring operations which are pertinent to precision agriculture. In this study, data mining operations and analysis are tested towards differentiation between vegetation and non-vegetation as the first step towards estimation of cropping regimes or patterns using Moderate Resolution Imaging Spectroradiometer (MODIS) data. MODIS daily NDVI Satellite Image Time Series (SITS) undergoes a process of cloud filtering via a Temporal Window Operation and smoothing using a Fast Fourier Transform (FFT). A process of discretization and dimensionality reduction using symbolic representation of time series data method, Symbolic Aggregate approXimation (SAX) is implemented and the resulting strings are clustered in order to reveal information inherent in the temporal profiles. Vegetation and non-vegetation land cover types were classified using clustering of state sequences and state sequence transitions with accuracies of 59% and 62% respectively.

Keywords*—Cropping Regimes, Moderate Resolution, SITS, Clustering, SAX, Data Mining*

I. INTRODUCTION

REGIONAL food security is threatened by uncertain global climatic conditions and global commodity price fluctuations which have resulted in decreased yields and dependence on local food production respectively[1]. Developing countries are particularly vulnerable due to reliance on agriculture for individual and national revenue aggravated by diverse agroclimatic characteristics, growing economies, diverse agricultural production systems, and farm typologies [2] , [3].

To counter these challenges, adaptation strategies are imperative and are aimed towards achieving food security as defined by The Food and Agriculture Organization (FAO) [4]. Adaptation and mitigation strategies include but are not limited to, formulation of short and long-term policies for improvement, sustenance, and protection of natural resources as in [3], modification of farming practices via technological uptake and adaptation of new crops and cropping systems [5]. There is therefore an exigency for timely and dependable information on agricultural production for capacity building, forecasting and constitution of contingency plans for vulnerable areas [3], [5] and [6].

To this end, identification and mapping of crop types is necessary since cropland maps, used in concert with meteorological information, enable monitoring of crop growth conditions and their response to climatic variations. The temporal frequency of this cropland information is therefore a key consideration, with the ideal being near real-time, as is the case with meteorological information.

Cropland mapping has been extensively researched utilizing a huge array of data and methods [7]-[12]. A vast majority of the studies implement supervised classification techniques such as maximum likelihood, nearest neighbours, logistic regression, decision tree algorithms, spectral unmixing methods and neural network methods, to list a few, which require in-situ observations and expert intervention for interpretation thus making them time consuming and expensive [5]. In addition, these methods are limited in their application to acquisition of near real-time information at variable spatial scales.

Moderate-resolution remote-sensing satellite systems such as MODIS provide data at regular high temporal frequency thus making them a valuable source of information for the estimation of crop area and type [13]. However, due to land cover and land use spatio-temporal heterogeneity and loss of data due to sensor and atmospheric artefacts the application of traditional hard classification approaches to MODIS data may result in significant errors in crop area estimation [12]. As such, SITS analysis has emerged as a powerful and useful tool for land cover classification and obtaining the baseline cropland information necessary for crop mapping and monitoring [13]-[17]. While SITS analysis provides information at a high temporal frequency, the large volume of data used in the analysis makes it computationally expensive thus limiting its

E. Nduati, Center for Environmental Remote Sensing (CEReS), Chiba University, Japan; (e-mail: nduatie@chiba-u.jp).

application to large spatial scales. Recent studies have therefore focused on exploitation of data mining techniques formerly native to sectors such as finance, telecommunications and biology for applications in remote sensing information extraction [13]-[22].

The abundance of data provided by many sensors at a variety of spatial and temporal resolutions, coupled with the need for powerful data analysis tools, has been described as a *data rich but information poor* situation [17], [23]. Data mining is one of the essential steps in the Knowledge Discovery Process and refers to discovery and extraction of information and knowledge from large amounts of data, usually gathered automatically [19], [23]. [24] Elucidate on the challenges pertaining to SITS analysis and the methods proposed in [13], [21] and [22] successfully use both the spatial and temporal dimension via sequential pattern mining for both high and moderate spatial resolution images. The Knowledge Discovery process as a whole involves: Data cleaning, Data integration, Data selection, Data transformation, **Data mining,** Pattern evaluation and Knowledge presentation [23].

In the context of remote sensing, implementing a Knowledge Discovery system for any spatial application using MODIS SITS data is a nontrivial task requiring significant computational power, a proper understanding of the various data products available and how they can be used towards the application's objective. There are over 80 MODIS data products available for scientific analysis, broadly categorized into Radiation Budget Variables, Ecosystem Variables and Land Cover Characteristics [25]. These datasets have varying spatial, temporal and spectral characteristics and therefore require diverse data handling procedures prior to analysis depending on the application.

Peri-urban cropland areas tend to have parcels of land whose area is less than the resolution of the highest native resolution MODIS product of 250m. As such, the issue of mixed pixels arises hence the need to first distinguish between vegetation and non-vegetation. Further, given that crop production is dynamic, a pixel identified as vegetation during the growing season will be bare land after harvesting within the same year, depending on the cultivated crop. This introduces pseudo-periodic behaviour as highlighted in [24]. For this reason, conventional LULC mapping which involves classification of single images as a snap shot in time would result in misclassification of uncultivated or fallow croplands.

The primary objective of this study is to develop a reproducible and extensible knowledge discovery framework and to demonstrate the implementation of machine learning algorithms towards estimation of cropping areas and patterns using a MODIS daily year-long data set. Vegetation and Non-vegetation areas are distinguished based solely on the temporal evolution of NDVI.

II. MATERIALS AND METHODS

A. Study Area

The proposed methodology is tested on Tomisato Shi, a cropland area of approximately 53.8 Km^2 located in the latitudinal range of N35 ^{0} 46'37" and N35 ^{0} 40'30" and longitudinal range of E140 ^{0} 16'44' and E140 ^{0} 24'31", in Northern Chiba Prefecture, Japan as shown in Fig. 1. The area is characterized by mixed land-use including urban land-use consisting of urban infrastructure such as roads, and residential and commercial buildings as well as agricultural land-use made up of rice paddy fields and small-holder horticultural crop farms.

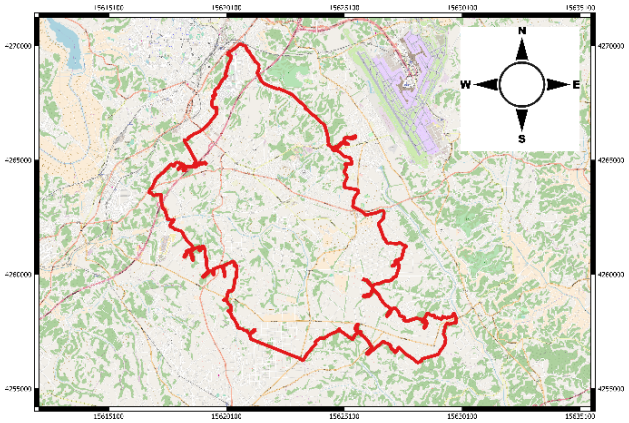


Fig. 1: Study Area of Tomisato Shi

The area can be considered to be a peri-urban food production region due to its close proximity to Tokyo and being located south-west of Narita international airport a busy international transportation hub. The main crops include rice, *suika* (watermelons), peanuts, *ninjin* (carrots), and *satoimo* (taro). Other crops produced include green leafy vegetables such as *hourensou* (spinach) and *komatsuna* (Japanese mustard spinach), and root vegetables and spices such as *daikon* (radish), *jagaimo* (potatoes) and *shouga* (ginger). Tomisato shi's weather patterns are as depicted in Fig. 2 below.

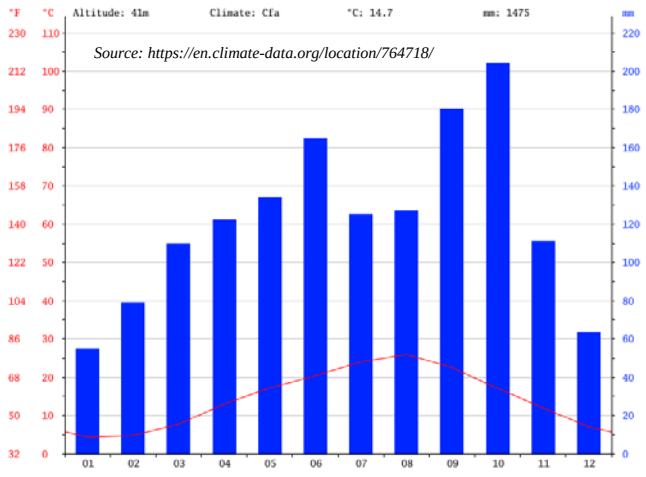


Fig. 2: Tomisato Shi climatograph

While the area receives adequate rainfall and temperatures allow for open field cultivation between April and November, farmers have the option to irrigate when necessary and each field has a water outlet installed. In addition, crops such as tomatoes and strawberries are mostly cultivated in greenhouses

in order to control the growth environment and protect them from pathogenic elements. Annual production of Japan's staple food -rice, compares favorably to the world's annual production. In 2016, Japan's rice production was 5.3 Tonnes/Ha while the world average was 3.1 Tonnes/Ha [26]. In contrast, production of horticultural crops such as soybean is relatively lower compared to world production figures which necessitates importation of complementary food crops. Further, with an ever increasing aging population, agricultural food production faces the challenge of abandoned croplands when farmers are no longer able to cultivate. To address these challenges, continuous monitoring of crop production in the spatial dimension is necessary.

B. Data

The data utilized in this study is the 250m daily MODIS surface reflectance product from 1st April 2016 to 31st March 2017. This product has two spectral bands, Red band (620~670nm) and NIR band (841~876nm) which are used to compute the Normalized Difference Vegetation Index (NDVI), which is a commonly used vegetation status and monitoring metric computed as in (1), [27].

$$\text{NDVI} = \frac{(NIR-Red)}{(NIR+Red)} \qquad (1)$$

The surface reflectance bands underwent pre-processing operations including band extraction from HDF scientific data set, reprojection, pixel-wise quality assessment, cloud-masking and gap-filling as illustrated in Fig. 3.

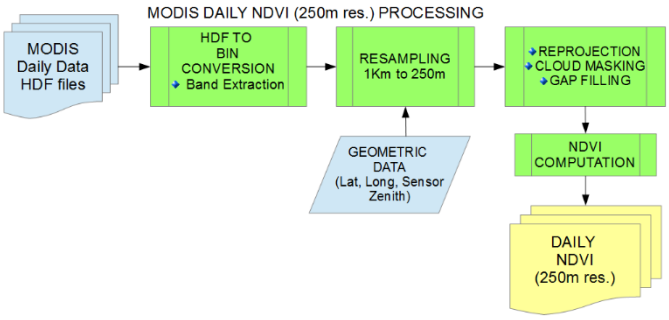


Fig. 3: MODIS data preprocessing workflow

The cloud-masking and gap-filling operations involved the use of a Temporal Window Operation and Fast Fourier Transform (FFT) to smooth and filter the NDVI signal for each pixel and thus have a continuous dataset with no missing values.

C. Validation Data Set

For this study, two sources of validation data were chosen. A high resolution Land Use/Land Cover (LULC) map generated from aerial photograph images for the region constituting the Inbanuma Basin, of which Tomisato Shi is a part of, and Google Earth images coinciding with the year of study, 2016/17. The LULC map was generated in 2007 and while the period of study is a decade later, it was deemed necessary to utilize this data in order to mimic scenarios where there is insufficient or old reference data, a challenged highlighted in [24].

D. Methods

The Daily 250m NDVI images were stacked to obtain a four-dimensional raster brick, hereafter referred to as MODIS_TS. For each pixel in MODIS_TS, the raw NDVI values are indexed by date and the pixel can thus be defined as:

$$\text{P(x, y)} = \{NDVI_{t1}, NDVI_{t2}, \dots\dots\dots, NDVI_{tN}\} \qquad (2)$$

Pixel location description can be defined in a variety of ways including a selected vertex – Upper Left Corner, Lower Right Corner – or the centroid, with most geospatial analysis software employing the former more commonly. In this study, since the intention was to relate the pixel values and hence the temporal evolution of NDVI to the land cover associated with that pixel with respect to the LULC map and high resolution google earth images, the LULC map was converted to a spatial grid object with a resolution matching the MODIS_TS. In every 250-meter grid there were several land cover types, thus the percentage cover of each type was computed and the centroid of the cover type with the highest percentage taken to be the pixel coordinate. These points were then used to extract the NDVI values across the temporal dimension from MODIS_TS resulting in an array (2) depicting the pixel-wise spatio-temporal evolutions.

$$Array \ni \begin{bmatrix} NDVI_{t_1}^{P(x_1,y_1)} & \cdots & NDVI_{t_N}^{P(x_1,y_1)} \\ \vdots & \ddots & \vdots \\ NDVI_{t_1}^{P(x_M,y_M)} & \cdots & NDVI_{t_N}^{P(x_M,y_M)} \end{bmatrix} \qquad (3)$$

The basis of time series analysis is the comparison of similarity or dissimilarity between data or observations at coincident temporal points by using distances, [24]. The choice of similarity/dissimilarity metric to use is therefore an important consideration and is greatly influenced by the indexing structure inherent in the data. Regardless of the geographical extent of the study area and the temporal frequency of observations, the high dimensionality of time series data impinges on the efficiency of the indexing structure and consequently, the suitability of similarity/dissimilarity measures to use for a given application. It is commonly referred to as the "dimensionality curse", which must be addressed via dimensionality reduction strategies which include but are not limited to, Discrete Fourier Transform (DFT), Adaptive Piecewise Constant Approximation (APCA), Piecewise Aggregate approXimation and Symbolic Aggregate Approximation (SAX) [28] – [35]. In this study, we sought to implement a dimensionality reduction method that preserved information on the temporal evolution of NDVI per pixel to the highest possible degree, while at the same time allowing for efficient and flexible handling of disparate volumes of time series data. SAX was chosen since it allows for discretization of the NDVI temporal evolutions into equiprobability symbolic representations that can then be efficiently used in data mining algorithms while also allowing for reverse mapping for validation purposes [36]. The row-wise values of the array derived from MODIS_TS were transformed into symbolic letter sequences of lengths 6 ("a" ~ "f") and 12 ("a" ~ "l") as described in [35] and [36]. It is worth noting that prior to discretization, *z*-normalization of the data set was carried out in

order to ensure that structural rather than amplitude driven similarities/dissimilarities remained the focus of the analysis and data mining operations [37]. Fig. 4 below shows the temporal profiles of the raw NDVI values for five sample points representing different land cover types in the study area.

Though the NDVI levels (amplitude) of the profiles are different especially between April and August, they exhibit similar structural behaviour. From September, some structural dissimilarities become apparent, particularly for profiles 107, 115 and 246. However, profiles 10 and 117 vary only in amplitude but remain structurally similar. In the context of data mining, the two profiles would be considered to be representative of the same land cover type or feature which may not be necessarily true as can be seen in Fig. 5.

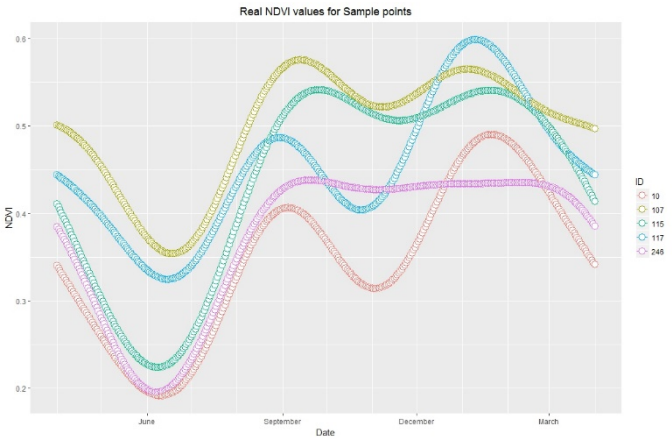


Fig. 4: Raw NDVI values

Following z-normalization, the structural similarities inherent in the months of April to August hold true while the structural dissimilarities in the subsequent months are amplified. This enables differentiation between distinctly different profiles based solely on structure and discrimination between profiles that have subtle differences based on time of occurrence of peaks and valleys as seen for profiles 10 and 117, where profile 117 peaks and valleys occur earlier than those of 10.

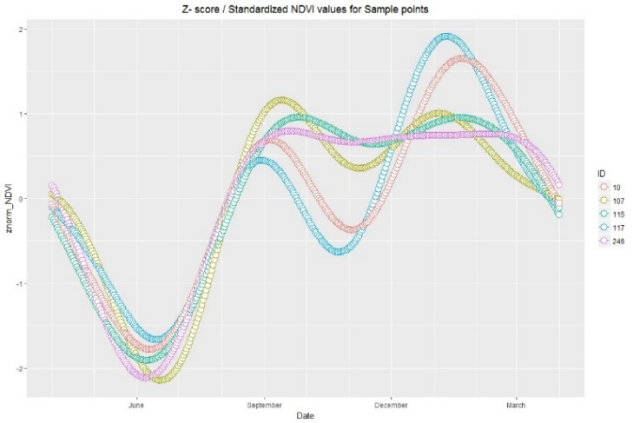


Fig. 5: z-Normalized NDVI

III. RESULTS AND DISCUSSION

The lengths of the SAX alphabets tested in this study of 6 and 12, were chosen in order to capture a pixels NDVI behaviour at bi-monthly and monthly rates. As one of the future objectives of this work is extensibility via use of multi-satellite data that have different temporal resolution, assessing the daily data in a manner that allows for introduction of finer spatial resolution data with lower temporal resolution could be beneficial. Fig. 6 is a plot of the temporal profile of a forest sample point under SAX discretization of a 6-letter alphabet. The resulting string sequences for each of the pixels in MODIS_TS carry information on the level of NDVI with respect to the bi-monthly divisions. In the case of the sample point depicted in Fig. 6, NDVI is in decline from DOY 0 to DOY 092 where it bottoms out and thereafter rises to the maximum within approximately one month. It then remains relatively high and stable until DOY 283 from whence it declines.

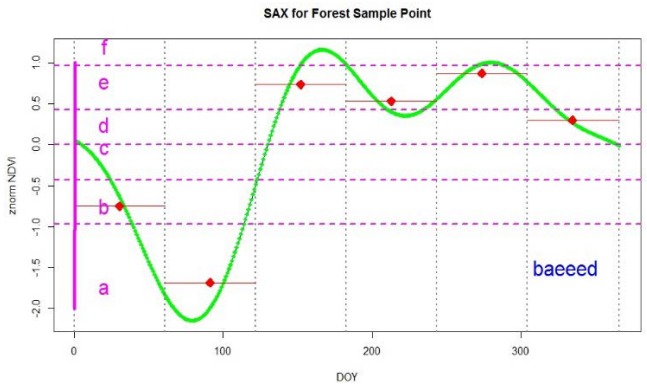


Fig. 6: SAX 6-letter discretization for sample point

The string sequences can therefore be seen as being representative of the states of the feature as a function of the feature type and its response to external conditions that result in changes to NDVI. Similarities/Dissimilarities between or among string sequences were therefore used in an unsupervised clustering algorithm to determine the natural groupings inherent in the data that can be related to the feature type or land cover type, [38], [39]. Though the state sequence mining methods in [38] were developed for Social Science application, in this study we show that they can be utilized for knowledge discovery for spatio-temporal data. Two approaches were implemented. The first approach used the 6-letter sequences labelled with aliases based on an ordinal scaling of the z-normalized NDVI into: “a” = “Very Low”, “b” = “Low”, “c” = “Moderate”, “d” = “Moderately High”, “e” = “High” and “f” = “Very High”. The second approach transformed the 12-letter sequence into state transitions by differencing, resulting in 11 transitions. Negative differences were classified as “Decay”, positive differences as “Growth” and zero or non-changing transitions as “Latency”.

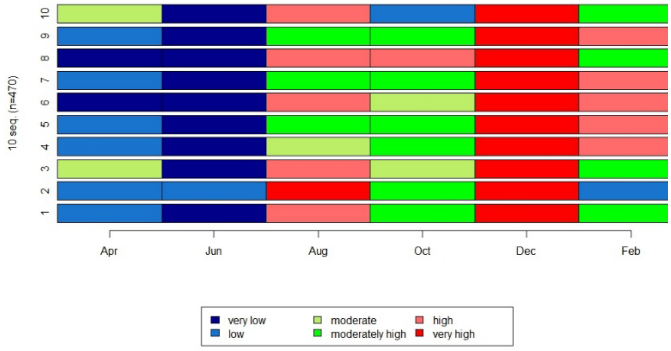


Fig. 7: Index distribution of 6-letter state sequences

The figure above shows the ordinal sequence of NDVI based on the 6-letter string sequence of 10 sample points. This graphic reveals that for the sample points, there is great similarity in two periods: June ~ July and December ~ January. The period between August and November shows the greatest variability and can thus be exploited for dissimilarity measures. This in essence tells us where to look for information and further gives us an idea of how the feature space is behaving within a short period of time. In Japan, the four seasons are clearly defined and while looking at spatio-temporal data in this perspective, the robustness of data mining algorithms in deriving phenological metrics from real NDVI SITS data is realized since NDVI variations as a result of seasonal changes can be distinguished from variations as a result of changes in feature characteristics given a sufficiently long time series dataset. In the spring and summer seasons as vegetation moves from dormancy and flourishes, the recorded reflected radiation from the earth's surface will change accordingly. However, given that this is a short time series, seasonality on account of atmospheric conditions cannot be completely ruled out.

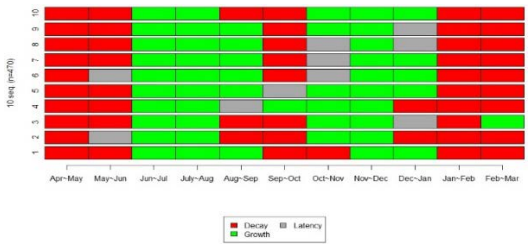


Fig. 8: Index distribution for 12-letter state transition sequences

The 11 state transitions obtained from differencing over the monthly 12-letter sequences are shown in Fig. 8 above. The state transitions provide a perspicuous representation of the temporality and order of states where, while in Fig. 7 the ordinal NDVI state sequence is predominantly "very low" in June to July, here we see that there is actually growth for all points but the temporality and duration varies.

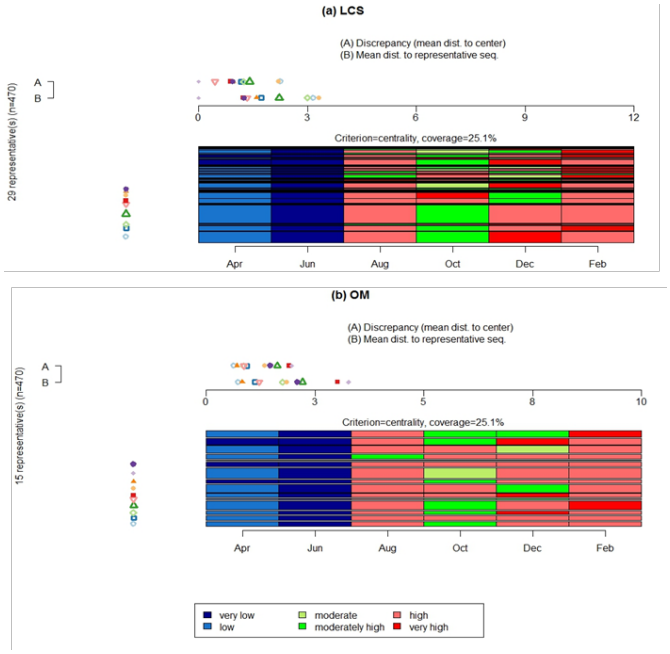


Fig. 9: State sequence LCS and OM clustering representative groups

In the December to January period where the ordinal scale for all 10 sample points is "Very high", we see that points 3, 8 and 9 are actually dormant and points 2 and 4 are in decline. This highlights the power of data mining in knowledge discovery and by utilizing techniques typically employed in the social sciences enhances transdisciplinarity.

Unsupervised classification or clustering is a useful way of exploring time series data and gaining insight into the natural groupings inherent in it and is based on dissimilarity measures. Dissimilarity measures of sequence should account for the order of the states and transitions in each sequence, the temporality of transitions and the duration of stay in each sequence [39]. Longest Common Subsequence (LCS) and Optimal Matching (OM) were the two dissimilarity measures used in an agglomerative hierarchical clustering algorithm. From the validation data sets, the pixels considered were generalized to the two main classes of interest – vegetation and non-vegetation. There were 470 pixels and therefore string sequences of an equal number of which, 186 were unique for the 6-letter string sequences. An ideal clustering method should yield comprehensible structures and this can be evaluated on the basis of the expected number of representative classes or groups. For the 6-letter sequences, LCS yielded 29 representative classes while OM yielded 15. Given that the original LULC map from which validation data was derived has 8 classes, the OM outperformed LCS but the number of representative classes is still far too large to yield any sensible structures. This may be indicative of hidden patterns which if accounted for would reduce this number further.

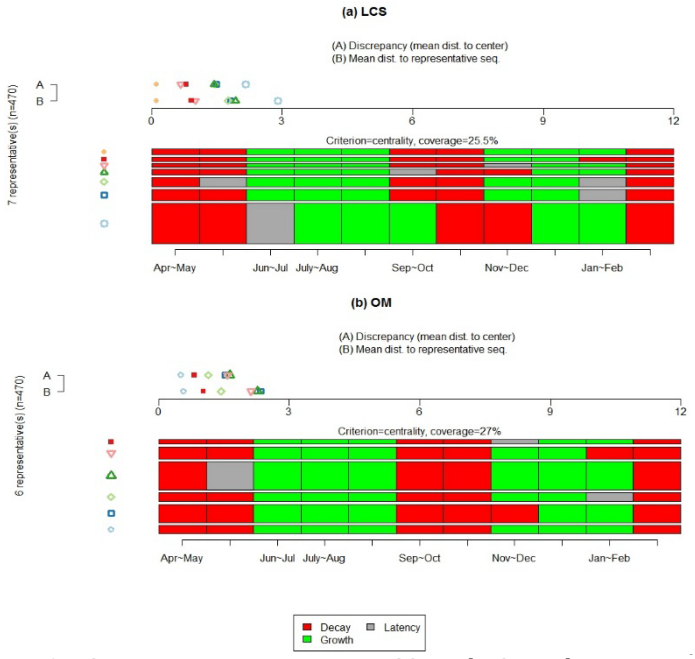


Fig. 10: State sequence transition LCS and OM clustering representative groups

Fig. 10 above shows the results of the LCS and OM dissimilarity measures for the state sequence transitions. The performance of OM surpasses that of LCS in this case as well. However, for both LCS and OM, the representative groups are greatly reduced to 7 and 6 respectively, with 153 unique sequences. This indicates that when looking at the state sequences at a finer temporal scale, and then obtaining the state transitions, the model becomes more intuitive. This is further demonstrated by the results of the agglomerative hierarchical clustering floored to two main classes of interest. Tables I and II are the confusion matrices as a result of clustering to two classes with “Class 1” and “Class 2” in the validation datasets representing vegetation and non-vegetation respectively. The state sequence clustering based on the 6-letter alphabet has an accuracy of 59% while the state sequence transition clustering based on 12-letter alphabet sequence differencing has an accuracy of 62%. These accuracies are deemed to be satisfactory in view of the fact that the data set is imbalanced since 84% of the points in the validation datasets when generalized to two main classes were vegetation.

Table I: State sequence clustering confusion matrix

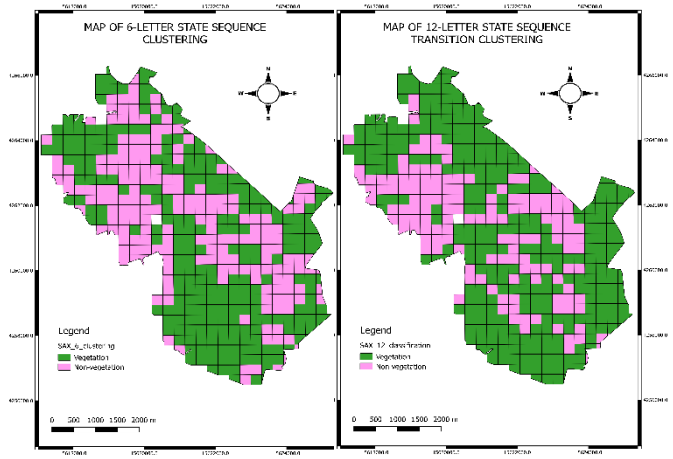
Actual / Predicted	Class 1	Class 2
Class 1	238	161
Class 2	34	37

Table II: State sequence transition clustering confusion matrix

Actual / Predicted	1	2
1	272	127
2	52	19

In order to fulfill the condition for reverse mapping and thus ensuring that the results of this study can be represented in the

patial context, the predicted classes were fed back to a GIS nvironment and are visualized as below in a grid spatial map f Tomisato Shi, Fig. 11. The validation dataset is also shown ı Fig. 12.



g. 11: Spatial grid maps of state sequence and state sequence ınsition clustering

From the maps above, the 6-letter state sequence clustering over estimates non-vegetation, though there are areas consistent with the 12-letter state sequence transition clustering. The underlying reasons for these disparities were not identified in this part of the study. Those pixels identified as vegetation with the greatest accuracy, that is the OM clustering method, are then used in the next stage of the study to discriminate between cropland and non-cropland regions and hence estimate the cropping regimes or patterns in the cropland areas.

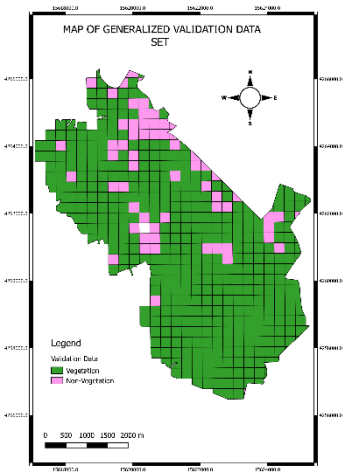


Fig. 12: Spatial grid map of validation data set

IV. Conclusions

In this study, we demonstrate the application of data mining techniques to differentiation between vegetation and non-vegetation land cover types using a high temporal frequency, moderate spatial resolution SITS dataset. In the absence of up-to-date reference or validation datasets, SITS analysis and unsupervised classification methods can go a long way in providing near-real time information necessary for agricultural monitoring and crop inventory mapping.

The results of this study illustrate the need for *z*-normalization of SITS data. Further, by discretizing the temporal profiles into string sequences, we were able to differentiate between vegetation and non-vegetation using an agglomerative hierarchical unsupervised classification method with satisfactory accuracies. Transforming the string sequences from states to state transition sequences enhanced the accuracy of the dissimilarity measures by decreasing the number of unique string sequences and the representative samples to yield intelligible groupings that can be related to land cover types.

All algorithms and processes were implemented using open source software, R 3.4.3 [40] and QGIS [41], demonstrating their immense capacity. Sequence mining was implemented using package TraMineR [38] which has been extensively used in social sciences. We demonstrate in this study how to apply the sequence mining methods designed for social science to spatio-temporal data.

ACKNOWLEDGEMENT

E. Nduati wishes to express deep appreciation to the Japan International Cooperation Agency (JICA) and the Africa-ai-Japan project for funding this research.

REFERENCES

[1] Brown, M. E., & Funk, C. C. (2008). Food security under climate change.

[2] Lobell, D. B., Burke, M. B., Tebaldi, C., Mastrandrea, M. D., Falcon, W. P., & Naylor, R. L. (2008). Prioritizing climate change adaptation needs for food security in 2030. *Science*, *319*(5863), 607-610.B.

[3] Jat, M. L., Dagar, J. C., Sapkota, T. B., Govaerts, B., Ridaura, S. L., Saharawat, Y. S., ... & Stirling, C. (2016). Climate change and agriculture: Adaptation strategies and mitigation opportunities for food security in South Asia and Latin America. In *Advances in agronomy* (Vol. 137, pp. 127-235). Academic Press.

[4] FAO. 2004. Food and Agriculture Organization FAOSTAT Data. Rome, Italy.

[5] Waldner, F., Canto, G. S., & Defourny, P. (2015). Automated annual cropland mapping using knowledge-based temporal features. *ISPRS Journal of Photogrammetry and Remote Sensing*, *110*, 1-13.J. Wang, "Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication)," *IEEE J. Quantum Electron.*, submitted for publication.

[6] Toma, L., Barnes, A. P., Sutherland, L. A., Thomson, S., Burnett, F., & Mathews, K. (2016). Impact of information transfer on farmers' uptake of innovative crop technologies: a structural equation model applied to survey data. *The Journal of Technology Transfer*, 1-18.

[7] Waldner, F., Fritz, S., Di Gregorio, A., & Defourny, P. (2015). Mapping priorities to focus cropland mapping activities: Fitness assessment of existing global, regional and national cropland maps. *Remote Sensing*, *7*(6), 7959-7986.

[8] Matton, N., Canto, G. S., Waldner, F., Valero, S., Morin, D., Inglada, J., & Defourny, P. (2015). An automated method for annual cropland mapping along the season for various globally-distributed agrosystems using high spatial and temporal resolution time series. *Remote Sensing*, *7*(10), 13208-13232.

[9] Shao, Y., Lunetta, R. S., Ediriwickrema, J., & Iiames, J. (2010). Mapping cropland and major crop types across the Great Lakes Basin using MODIS-NDVI data. *Photogrammetric Engineering & Remote Sensing*, *76*(1), 73-84.

[10] Xiong, J., Thenkabail, P. S., Gumma, M. K., Teluguntla, P., Poehnelt, J., Congalton, R. G., & Thau, D. (2017). Automated cropland mapping of continental Africa using Google Earth Engine cloud computing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *126*, 225-244.

[11] Biradar, C. M., Thenkabail, P. S., Noojipady, P., Li, Y., Dheeravath, V., Turral, H., & Xiao, X. (2009). A global map of rainfed cropland areas (GMRCA) at the end of last millennium using remote sensing. *International journal of applied earth observation and geoinformation*, *11*(2), 114-129.

[12] Lobell, D. B., & Asner, G. P. (2004). Cropland distributions from temporal unmixing of MODIS data. *Remote Sensing of Environment*, *93*(3), 412-422.

[13] Vintrou, E., Ienco, D., Bégué, A., & Teisseire, M. (2013). Data mining, a promising tool for large-area cropland mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *6*(5), 2132-2138.

[14] Hentze, K., Thonfeld, F., & Menz, G. (2016). Evaluating Crop Area Mapping from MODIS Time-Series as an Assessment Tool for Zimbabwe's "Fast Track Land Reform Programme". *PloS one*, *11*(6), e0156630.

[15] Eerens, H., Haesen, D., Rembold, F., Urbano, F., Tote, C., & Bydekerke, L. (2014). Image time series processing for agriculture monitoring. *Environmental Modelling & Software*, *53*, 154-162.

[16] Gómez, C., White, J. C., & Wulder, M. A. (2016). Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *116*, 55-72.

[17] Petitjean, F., Kurtz, C., Passat, N., & Gançarski, P. (2012). Spatio-temporal reasoning for the classification of satellite image time series. *Pattern Recognition Letters*, *33*(13), 1805-1815.

[18] Datcu, M., Seidel, K., Pelizarri, A., Schroeder, M., Rehrauer, H., Palubinskas, G., & Walessa, M. (2000). Image information mining and remote sensing data interpretation. In *Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International* (Vol. 7, pp. 3057-3059). IEEE.

[19] Velickov, S., Solomatine, D. P., Yu, X., & Price, R. K. (2000, July). Application of data mining techniques for remote sensing image analysis. In *Proc. 4th Int. Conference on Hydroinformatics, USA*.

[20] Li, D., Wang, S., Yuan, H., & Li, D. (2016). Software and applications of spatial data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *6*(3), 84-114.

[21] Julea, A., Méger, N., Bolon, P., Rigotti, C., Doin, M. P., Lasserre, C., & Lazarescu, V. N. (2011). Unsupervised spatiotemporal mining of satellite image time series using grouped frequent sequential patterns. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(4), 1417-1430.

[22] Petitjean, F., Gançarski, P., Masseglia, F., & Forestier, G. (2010, September). Analysing satellite image time series by means of pattern mining. In *International Conference on Intelligent Data Engineering and Automated Learning* (pp. 45-52). Springer, Berlin, Heidelberg.

[23] Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques.

[24] Petitjean, F., Inglada, J., & Gançarski, P. (2012). Satellite image time series analysis under time warping. *IEEE transactions on geoscience and remote sensing*, *50*(8), 3081-3095.

[25] https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table

[26] https://data.oecd.org/agroutput/crop-production.htm

[27] Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*, 127-150.

[28] Verleysen, M., & François, D. (2005, June). The curse of dimensionality in data mining and time series prediction. In *International Work-Conference on Artificial Neural Networks* (pp. 758-770). Springer, Berlin, Heidelberg.

[29] Keogh, E., Chakrabarti, K., Pazzani, M., & Mehrotra, S. (2001). Dimensionality reduction for fast similarity search in large time series databases. *Knowledge and information Systems*, *3*(3), 263-286.

[30] Vidaurre, D., Rezek, I., Harrison, S. L., Smith, S. S., & Woolrich, M. (2014). Dimensionality reduction for time series data. *arXiv preprint arXiv:1406.3711*.

[31] Agrawal, R., Faloutsos, C., & Swami, A. (1993, October). Efficient similarity search in sequence databases. In *International Conference on Foundations of Data Organization and Algorithms* (pp. 69-84). Springer, Berlin, Heidelberg.

[32] Chan, K. P., & Fu, A. W. C. (1999, March). Efficient time series matching by wavelets. In *Data Engineering, 1999. Proceedings, 15th International Conference on* (pp. 126-133). IEEE.

[33] Keogh, E., Chakrabarti, K., Pazzani, M., & Mehrotra, S. (2001). Locally adaptive dimensionality reduction for indexing large time series databases. *ACM Sigmod Record*, *30*(2), 151-162.

[34] Keogh, E. J., & Pazzani, M. J. (2000, August). Scaling up dynamic time warping for datamining applications. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 285-289). ACM.

[35] Lin, J., Keogh, E., Lonardi, S., & Chiu, B. (2003, June). A symbolic representation of time series, with implications for streaming algorithms. In *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery* (pp. 2-11). ACM.

[36] Lin, J., Keogh, E., Wei, L., & Lonardi, S. (2007). Experiencing SAX: a novel symbolic representation of time series. *Data Mining and knowledge discovery*, *15*(2), 107-144.

[37] Goldin, D. Q., & Kanellakis, P. C. (1995, September). On similarity queries for time-series data: constraint specification and implementation. In International Conference on Principles and Practice of Constraint Programming (pp. 137-153). Springer, Berlin, Heidelberg.

[38] Gabadinho, A., Ritschard, G., Mueller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, *40*(4), 1-37.

[39] Studer, M., & Ritschard, G. (2016). What matters in differences between life trajectories: A comparative review of sequence dissimilarity measures. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *179*(2), 481-511.

[40] R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[41] "QGIS Development Team (YEAR). QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

Investigation of Surface Quality in High Speed Welding of Aluminum Using Adjustable Ring-mode Fiber Laser

M. R. Maina, Y. Okamoto, M. Närhi, J. Kangastupa, J. Vihinen and A. Okada

***Abstract*—**Sheet metals joint technology is highly required in many manufacturing industries where aluminum finds great application. Laser welding of aluminum has difficulties due to its low melting point, high thermal conductivity and low absorptivity. Upon reaching the melting point, the absorptivity increases with increasing temperature. In deep penetration welding, a lot of spatter arise hence deterioration of surface quality, while good surface quality is an essential factor to improve component functionality. In order to achieve deep penetration with stable welding phenomenon and ensure good surface quality, at a high welding speed, adjustable ring-mode fiber laser has been employed, since it offers high power with a dynamic adjustable beam profile consisting of a center part and a ring-part. Pure-ring-mode irradiation using ring power only and dual-mode irradiation using both center power and ring power, were investigated through experiments and numerical simulations. Dual-mode irradiation made it possible to stabilize the welding process. The center power helps to achieve sufficient deep penetration, while ring power ensures good temperature distribution.

***Keywords*—**Adjustable ring-mode fiber laser, Aluminum, Laser welding, Surface quality.

I. INTRODUCTION

LASER welding technology has been used in many industries because of its high accuracy and efficiency. In comparison with conventional welding processes, laser welding offers the benefits of precision control of heat input, minimum thermal distortion, and small heat affected zones (HAZ) and excellent repeatability [1]. In addition, deep penetration welding capability of lasers makes it possible to weld thick metal sheets and also offers a great potential in joining of materials such as aluminum.

In this study, a non-heat treatable, Al-Mg alloy 5022 GC45-O was used. The addition of magnesium to aluminum increases strength through solid solution strengthening and improves their strain hardening ability. Recent application technologies of aluminum to automobiles have shown that aluminum alloy 5022 finds great application in vehicle body panels, since it is a high-strength, high-formability and light weight material [2]. These applications demand for an efficient and high speed method of joining aluminum metal sheets, with the aim of achieving good surface quality and high productivity.

There is a limit to which welding speeds can be increased without affecting the weld bead quality. The humping phenomenon sets a limit to the processing speed in laser beam welding for a given laser beam intensity. In order to show humps formation at high welding speeds, Fabbro and Amara developed a 3-D transient model based on numerical resolution of the fluid flow and heat transfer equations for deep penetration laser welding at high welding speed condition, which resulted in humping phenomenon [3]. When this phenomenon occurs, the height of the weld bead varies with the welding speed [4]. For a given configuration of material and processing parameters, there arises melt pool instability upon exceeding a certain welding speed. Thomy et al., through numerical and experimental investigations using a single-mode fiber laser, have reported that at a high welding speed, melt flow around the keyhole may result in a stagnation area behind the keyhole, destabilizing the melt pool surface. This promotes the onset of humping, which is characterized by formation of periodic droplets and severe undercuts on the weld bead surface [5].

The prerequisite for the onset of humping is an upward oriented melt pool stream due to drag forces. The higher the speed, the smaller the inclination angle of the absorption front, thus the upward momentum of the melt pool increases. The smaller the focus diameter, the shorter the humping length. The drag forces are reduced as the evaporation pressure at the absorption front decreases with lower intensity [6].

Laser welding of aluminum has difficulties due to its low melting point, high reflectivity and high thermal conductivity. The ideal situation for laser welding is to have materials with high absorptivity and low thermal conductivity. Normally, the absorptivity varies depending on the temperature; as a material becomes hot, the absorptivity increases. For aluminum, upon reaching the melting point, the absorptivity increases with increasing temperature and the process becomes unstable. A lot of spatter arise, and deep penetration welding with good surface quality becomes a challenge. In addition, at high welding

M. R. Maina, Nontraditional Machining Laboratory, Okayama University, Japan (phone: +819022901655; fax: +81862518039; e-mail: maina@ntmlab.mech.okayama-u.ac.jp).
Y. Okamoto, Nontraditional Machining Laboratory, Okayama University, Japan (e-mail: Yasuhiro.Okamoto@okayama-u.ac.jp).
M. Närhi, Corelase Oy, Finland (e-mail: matti.narhi@corelase.fi)
J. Kangastupa, Corelase Oy, Finland (e-mail: jarno.kangastupa@corelase.fi)
J. Vihinen, Laser Application Laboratory, Tampere University of Technology, Finland (e-mail: jorma.vihinen@tut.fi)
A. Okada, Nontraditional Machining Laboratory, Okayama University, Japan (e-mail: okada@ntmlab.mech.okayama-u.ac.jp)

speeds, the humping phenomenon leads to deterioration of the surface quality.

There is need for deep penetration welding, and also maintenance of good surface quality for the various applications of aluminum, in order to enhance component functionality. Nakashiba et al. have shown that, for good surface quality, boundary region between keyhole and heat conduction welding mode would have advantages, but stabilization of the welding phenomenon is important [7].

Recently, the advancements of laser technologies have led to availability of lasers with high power and high beam quality, such as adjustable ring-mode (ARM) fiber lasers. ARM fiber laser is suitable for wide range of materials and thicknesses. It offers a dynamic adjustable beam profile with two parts namely; center part and ring part, hence a unique form of intensity distribution.

To achieve deep penetration with stable absorption and ensure good surface quality, at a high welding speed, ARM fiber laser has been employed. There is possibility to perform superior temperature distribution by combination of center power and ring power. Therefore, laser welding of aluminum, using ARM fiber laser, was investigated experimentally and numerically using a finite element method (FEM) based model. The influence of laser intensity distribution on the weld bead was discussed.

II. EXPERIMENTAL METHOD

Overlap welding of two aluminum alloy 5022 sheets, each 1.5 mm thick, was performed at a welding velocity of 15 m/min. The processing head was fixed, while the specimen was moved. The back side of the sheets was left open to air. The processing head employed concave mirrors with a collimation of 100 mm and a focal length of 150 mm. The laser wavelength was 1070 nm, and the beam spot diameters for the outer ring and center were 270 μm and 105 μm, respectively. The fiber used in this study had an inner core diameter of 70 μm, an outer core diameter of 180 μm and a gap of 15 μm, as shown in Fig. 1. The gap is filled with a glass material of low refractive index. Fig. 2 shows the intensity distributions for dual-mode and pure-ring-mode operations. The beam parameter products for the center and ring part were 2 mm·mrad and 8 mm·mrad, respectively.

Nitrogen was used as the shielding gas. It was supplied from the backside, at a flow rate of 15 L/min, through an off-axial nozzle of size 7.0 ×1.5 mm, inclined at an angle of 45° to the vertical. The gap between the off-axial nozzle tip and the workpiece surface was set at 10 mm. The experimental setup is schematically shown in Fig. 3. Table 1 shows the combinations of laser center power and ring power used in this study.

Weld bead evaluation involved measurements of bead width and height, roughness measurement, evaluation of surface appearance, and observation of cross-section. A laser displacement sensor (LDS) with a repeatability of 0.2 μm was used for weld bead profile measurement as shown in Fig. 4. Weld bead surfaces were observed using an optical microscope. Measurement of roughness involved use of stylus type roughness measurement machine with a stylus radius of 2 μm. For cross-section observation, the surface was first polished and then etched using HCl with 35% weight concentration.

TABLE 1
LASER POWER COMBINATIONS.

Case	Center power P_c (kW)	Ring power P_r (kW)	Total power (kW)
A	0	5.0	5.0
B	1.5	3.5	5.0

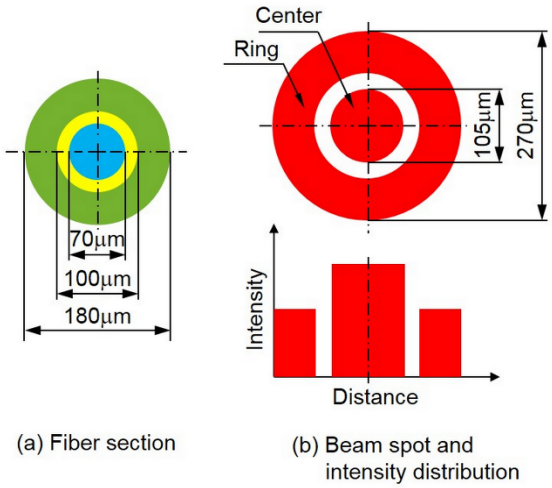


Fig. 1 Schematic illustration of fiber section, beam spot and intensity distribution.

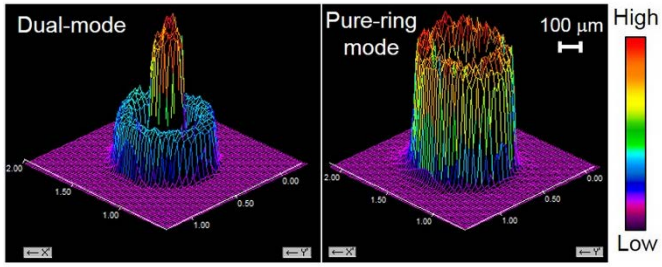


Fig. 2 Intensity distributions for dual-mode and pure-ring-mode operations.

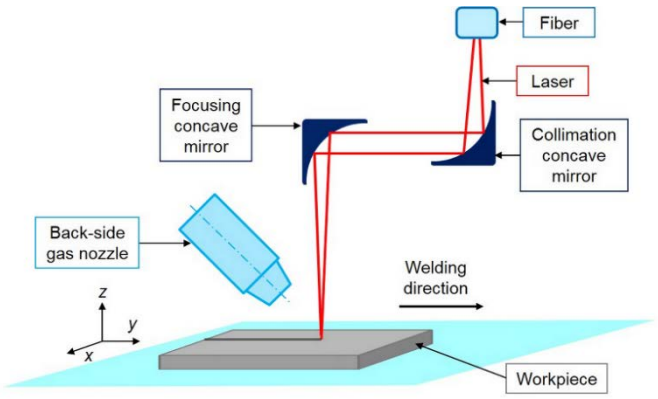


Fig. 3. Schematic illustration of experimental setup.

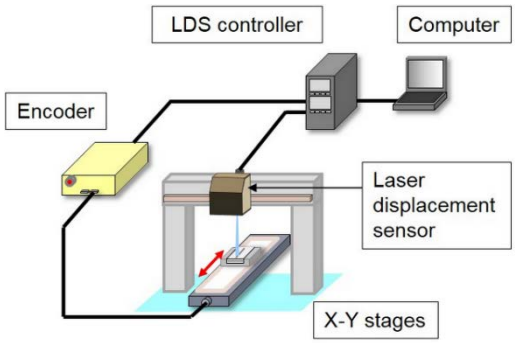


Fig. 4 Setup for weld bead profile measurement using laser displacement sensor.

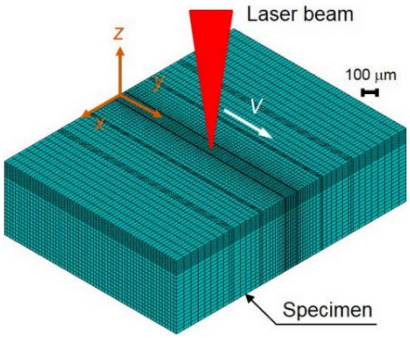


Fig. 5 FEM model for laser welding process.

III. NUMERICAL ANALYSIS METHOD

To further study the welding phenomenon in welding of aluminum using ARM laser, numerical simulation was performed using the general finite element program 'ANSYS'. Center and ring power combinations similar to experimental work were used. The geometry model employed similar total thickness of 3 mm as the experimental work. However, a small section of 8 mm length and 6 mm width was chosen. The transient temperature distribution of the welded specimen was a function of time *t* and Cartesian coordinate system with *y*-axis along the welding direction, *z*-axis along the thickness direction, and the origin located on the specimen surface as shown in Fig. 5. The governing equation for the transient heat conduction can be written as:

$$\rho c \frac{\partial T}{\partial t} = k\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}\right) + Q_i \quad (1)$$

where, T is the temperature, ρ the material density, c the specific heat, k the thermal conductivity, and Q_i the internal heat generation per unit volume.

The temperature dependent material properties are important for the accurate calculation of temperature field. The temperature response in a material involved in high heat fluxes is determined by thermal material properties of thermal conductivity, specific heat and density, which are dependent on temperatures [8]. Enthalpy values around the fusion and vaporization regions were computed in order to take into consideration the latent heat of aluminum. Aluminum has a melting point of 933 K and an evaporation point of 2741 K.

The welding process was performed in the middle of the specimen. Meshes were graded such that they were finest in the region of highest and most rapid temperature gradient near the heat input. Course meshes were used for the regions farthest from the heating zones. In addition, element sizes increased across the thickness of the specimen, being finer near the heated side of the specimen. The initial condition for the entire region was set at a room temperature of 293 K. The heat flux generated by the laser beam was applied on the top surface of the specimen. For the non-irradiated top surface and *x*-*z* end planes, the heat flux was assumed to be convective with a heat transfer coefficient of 10 W/(m²·K), while *y*-*z* end planes had infinite boundary temperature. The heat source was moved along *y*-axis in steps of 0.08 ms at a scanning velocity of 250 mm/s, with beam spot diameters of 270 µm in the outer ring and 105 µm at the center.

The absorption rate of a material is related to the material resistivity and the wavelength of the laser irradiation. However, absorptivity increases with increase in keyhole depth. Surface absorptivity is affected by mechanisms related to surface geometry during keyhole generation, through creation of surface ripples, which enhances absorptivity [9]. In this study, the absorptivity of aluminum at the top surface was taken to be 0.2 from literature [10], but increased gradually with heat input across the thickness. The FEM model used a combination of surface heat source and volumetric heat source. The total heat input was computed from the summation of surface heat source on the top surface and the volumetric heat source along the thickness direction as shown in equation 2.

$$Q = Q_s + Q_v \quad (2)$$

where Q is total heat input, Q_s surface heat flux, and Q_v volumetric heat flux. It was assumed that 25% of the heat power was absorbed on the surface of the specimen and the remaining 75% was absorbed by the keyhole wall. The surface heat flux was applied in top-hat mode, while the volumetric heat flux was applied in Gaussian distribution as expressed by equations 3 and 4, respectively.

$$Q_s = P_s/(\pi R^2) \quad (3)$$

$$Q_v = \frac{3P_v}{\pi {r_0}^2 h_d} exp\left(-3\frac{{r_c}^2}{{r_0}^2}\right)\cdot\left(1-\frac{z_i}{h_d}\right) \quad (4)$$

where R is beam spot radius, P_s power absorbed on the surface, P_v power absorbed by keyhole wall, r_0 initial radius of keyhole, h_d maximum keyhole depth, r_c current keyhole radius, and z_i current keyhole depth.

IV. RESULTS AND DISCUSSION

A. *Effects of laser power density on weld bead*

Fig. 6 shows weld bead surface and cross-section appearances by using different intensity distributions under the same total power condition, while Fig. 7 shows the

corresponding bar graph for variations of weld bead height, penetration depth and surface roughness. Both cases of pure-ring-mode, where only ring power is used, and dual-mode, where both center power and ring power are used, showed almost equal penetration depths at the same total power of 5.0 kW. However, weld bead roughness was considerably higher in the case of pure-ring-mode. Center power helps to initiate faster keyhole formation and a small keyhole diameter is expected. This helps to reduce spattering and prevent humping, hence reduction in surface roughness of weld bead. However, too much center power may lead to high material removal, hence deterioration of the weld bead.

The cross-sectional profile of the weld bead for dual-mode welding presents an 'hourglass' shape, which means that the top is the widest, the bottom narrower, and the middle the narrowest. This can be explained by influence of surface tension gradients in the molten metal flow. The maximum flow velocity of molten metal appears on the surface of melt pool, where there is the maximum temperature gradient. The molten metal flows from the higher-temperature keyhole boundary at the center of the beam to the lower temperature melt pool boundary. This leads to an expansion of the fusion zone in the top and bottom surface areas. For pure-ring-mode welding, since only ring power is present, a larger keyhole diameter is expected, and thus

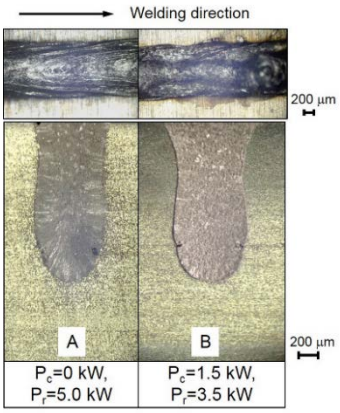


Fig. 6 Surface and cross-section appearance under different power densities.

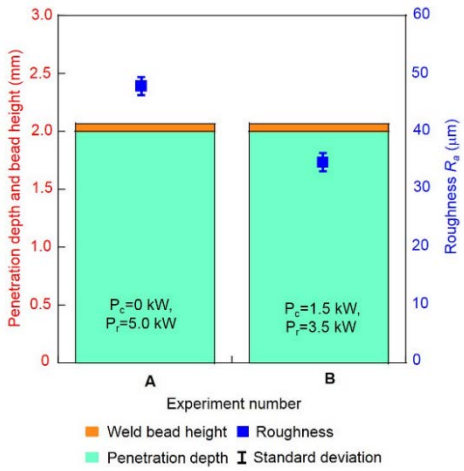


Fig. 7 Variation of weld bead height, penetration depth and surface roughness under different power densities.

the surface tension gradients in the molten metal flow is low. It is considered that heat is uniformly distributed around the ring, hence the uniform profile.

B. Influence of temperature distribution

The weld bead profiles obtained from experiments and simulations, under similar laser processing parameters, were compared in order to verify the simulation results. The simulation results gave a good estimation of the weld bead cross-section. Fig. 8 shows the temperature distribution isotherms on the top surface for dual-mode and pure-ring-mode cases where 5.0 kW total power was used in each case. The red part represents the molten and evaporated material at the center for temperatures above the melting point 933 K. The weld pool shapes around the high-energy heat source appear different. The 'tail' in dual mode case presents a bigger radius than that in pure ring mode case. The 'tail' shows the behavior of heat flow as the heat source moves along the surface. Dual mode case leads to a slower cooling rate compared to that of pure ring mode case.

Center power leads to faster keyhole formation thereby enhancing deep penetration. With sufficient keyhole, spattering is reduced and there is stability of the surface state of molten metal. A smaller keyhole diameter in dual mode case means that the melt pool region between the keyhole and the solid material is wider leading to a smooth flow. On the contrary, for pure ring mode, there is a larger keyhole diameter and the keyhole depth is smaller as shown in Fig. 9.

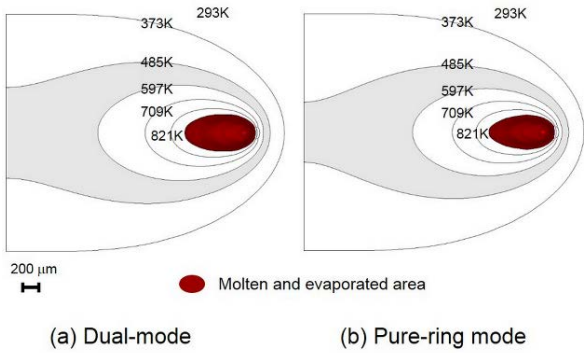


Fig. 8 Temperature distribution isotherms on the top surface for dual-mode and pure-ring-mode welding.

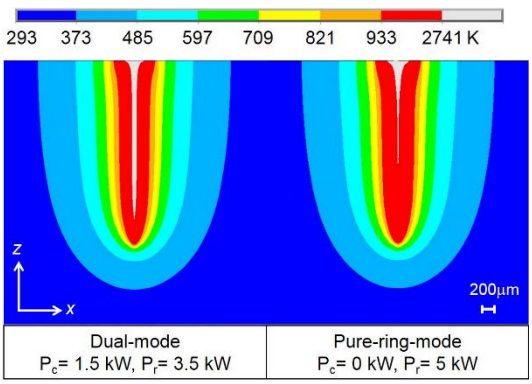


Fig. 9 Variations of keyholes for dual-mode and pure-ring-mode welding.

Despite having a smaller keyhole depth for pure-ring-mode welding, deep penetration could be achieved due to high power density at the ring. The temperature of the melt pool near the evaporated area was very high, being at the liquid-vapor transition region. From this study, dual-mode welding with high power ARM fiber laser has shown the capability of obtaining good surface quality and sufficient deep penetration, at high welding speeds.

V. CONCLUSIONS

Laser welding using adjustable ring-mode fiber laser proves to be a reliable high speed joining method for aluminum with the possibility of obtaining high quality welds. The main conclusions obtained in this study are as follows:

1) Dual-mode welding leads to a weld bead with lower surface roughness compared to pure-ring-mode welding.
2) Center power helps to initiate faster and sufficient keyhole formation, while ring power ensures good temperature distribution, hence stabilizing the welding process.

ACKNOWLEDGEMENT

The research work covered in this paper was partly supported by Japanese Ministry of Education, Culture, Sports, Science and Technology, and Jomo Kenyatta University of Agriculture and Technology, Kenya.

REFERENCES

[1] S. Josefine, C. Isabelle, and F. Alexander, “Laser welding process – a review of keyhole welding modelling,” Physics Procedia, vol. 78, pp. 182-191, 2015.
[2] I. Takashi, T. Kenji, Y. Hiroyuki, T. Yoshiki, and I. Tadayuki, “Wrought aluminum technologies for automobiles,” Kobelco Technology Review, vol. 26, pp. 55-62, 2005.
[3] R. Fabbro and E. Amara, “Modeling of humps formation during deep-penetration laser welding,” Applied Physics A: Materials Science and Processing, vol. 101, pp. 111-116, 2010.
[4] G. Hongping and S. Boris, “A practical use of humping effect in laser beam welding,” Journal of Laser Applications, vol. 23, pp. 012001, 2011
[5] C. Thomy, T. Seefeld, and F. Vollertsen, “Humping effect in welding of steel with single-mode fibre laser,” Welding in the World, vol. 52, pp. 9-18, 2008.
[6] O. Andreas, A. Patschger, and S. Michael, “Numerical and experimental investigations of humping phenomena in laser micro welding,” Physics Procedia, vol. 83, pp. 1415-1423, 2016.
[7] S. Nakashiba, Y. Okamoto, T. Sakagawa, M. Harada, and A. Okada, “The boundary of keyhole generation in micro-welding of aluminum alloy by pulsed Nd:YAG laser with superposition of continuous diode laser,” Journal of the Japan Society for Precision Engineering, vol. 80, pp. 419-424, 2014.
[8] N. Kethireddy, S. Ammiraju, G. Kalvala, and G. Nallacheruvu, “Temperature dependence of density and thermal expansion of wrought aluminum alloys 7041, 7075 and 7095 by gamma ray attenuation method,” Journal of Modern Physics, vol. 4, pp. 331-336, 2013.
[9] R. Fabbro, F. Coste, S. Slimani, and F. Briand, “Study of keyhole behaviour for full penetration Nd:YAG CW laser welding,” J. Phys. D: Appl. Phys., vol. 38, pp. 1881-1887, 2005.
[10] T. Forsman, A. Kaplan, J. Powell, and C. Magnusson, “Nd:YAG laser welding of aluminium; factors affecting absorptivity,” Lasers in Engineering, vol. 8, pp. 295-309, 1999.

Economic Dispatch of PV-Integrated Power System with Optimally Sized Battery Energy Storage System using Particle Swarm Optimization

Lonah Nyaboke Segera

Abstract— Rapid escalation of fuel prices, depletion of fossil fuel reserves and environmental concerns have compelled power system operators to incorporate the Renewable Energy (RE) resources for example solar in the energy mix to meet the demand. Although the renewable energy resources are valuable and cost effective, they are unpredictable in nature and are dependent on weather conditions. The market for solar energy has been expanding rapidly worldwide. Solar-Photovoltaic (PV) systems generally have considerable power variations, which include voltage fluctuations and frequency variations. The intermittent power generation of a solar farm can perturb the supply and demand balance of the whole power system. Therefore, mitigating the adverse effects on the grid from an intermittent PV source has become essential for increasing the penetration level of PV systems. In the power system operations planning, the economic load dispatch of thermal generating system is one of the most important problems. Recent global inclination towards the utilization of more and more renewable energy makes this problem important than ever. Efficient and reliable planning of power system with significant penetration of these resources brings challenges due to their fluctuating and uncertain characteristics. Energy storages are emerging as a predominant sector for renewable energy applications. Recently, Battery Energy Storage System (BESS) has become a promising solution to help PV integration, due to the flexible real power control of the batteries. This research aims at conducting the economic dispatch of thermal and PV system with battery storage. The sizing methodology is optimized using Particle Swarm Optimization algorithm to minimize the cost of investment and losses incurred by the system in form of peak load shaving. The proposed methodology is tested and validated on a standard IEEE 30 bus test system.

Keywords—Battery energy storage system, Economic Dispatch, Photovoltaic, Particle swarm optimization, Renewable energy.

I. INTRODUCTION

The global electrical energy consumption is rising thereby increasing the demand of power generation. The use of renewable energy resources has become necessary due to depletion of fossil fuels that were widely used traditionally. Among alternatives used for the generation of electricity are a number of unconventional sources including solar and wind energy. Technological and economic progress of efficient and reliable solar-photovoltaic (PV) panels as well as the concerns about environmental issues has contributed to large penetration of solar energy in the power system [1]. With further developments in the PV technology and lower manufacturing costs, the outlook is that the PV power will possess a larger share of electric power generation in the near future. Grid-connected PV is ranked as the fastest-growing power generation technology [2]. Although the PV installation costs are still high, PV generates pollution-free and very cost-effective power that relies on a free and abundant source of energy [3]. However, the integration of these renewable sources into the power system exhibits challenges mainly due to their natural intermittency and limited predictability.

The economic dispatch (ED) is the short-term determination of the optimal output of on-line power generation units, to meet the system load, at the lowest possible cost without violation of any operational limits of generation and transmission facilities. To obtain the solution of the ED problem, we find the best distribution of the electrical power output from the available generating units. Earlier to the common use of alternate sources of energy, the ED problem looked after only the conventional thermal power generators, which use non-renewable resources as fuels. Now days it has become necessary that there be an alternate method of generation apart from the conventional thermal energy power generation, and one of the sources that has gained popularity is the Solar-PV. In recent decades, PV held a good position as there is no operating cost that causes a reduction in the total cost when integrated with the conventional system. Thus, PV applications became more practical for the generation of the power. In the case of huge penetration of varying power sources such as PV, due to the weather conditions or the day and night phenomenon thermal units requires a serious operation pattern because the outputs from PV generators are affected by the change in the radiation of solar energy. Thus, the large-scale utilization of solar energy depends on the flexible operation of the thermal units. Moreover, PV units can be paired with energy storage elements like batteries, to stack up the excess power generated during off-peak hours and give it back as generation during peak hours. For a system to have economic operation, the PV/Battery commitment and dispatch must be optimal [4]. This paper presents a proposed Economic Dispatch of a PV-integrated power system with optimally sized Battery Energy Storage System using PSO. The paper is organized as follows: in Section 2, a brief discussion on solar energy is given whereas its characteristics and integration challenges to the power system operators in regards to Economic Dispatch whereas the different types energy storage systems are presented in Section 3. Economic Dispatch is discussed in Section 4 while PSO is introduced in Section 5. The optimal BESS sizing-using PSO is given in Section 6, the problem formulation is given in Section 7. Section 8 gives the proposed method for Economic Dispatch while in Section 9 conclusions are drawn.

Lonah Nyaboke Segera Department of Electrical and Electronic Engineering JKUAT (+2540703436714; e-mail lonahsegera@gmail.com).

II. SOLAR ENERGY

Renewable energy (RE) sources, including wind, solar and their hybrid systems, have become attractive options of providing energy globally due to low cost, no pollutant emission, energy security, easy accessibility and reduction fossil fuel consumption [1, 2, 3, 4]. The major applications of solar energy include solar thermal system, which converts solar energy to thermal energy, and photovoltaic (PV) system, which converts solar energy to electrical energy. In recent decades, PV held a good position as there is no operating cost, which causes a reduction in the total cost when integrated with the conventional system. Photovoltaic (PV) array, can be stand-alone installed for providing electricity in some remote areas or be connected to the grid for selling power generated. Because of instantaneous and unstable nature of solar energy, PV usually works with battery storage to provide continuous and stable power, by use of the PV-battery hybrid system. Solar energy is intermittent and it causes most of the undesired effects, such as voltage variations, real and reactive power fluctuations and potential effects for overcurrent and overvoltage protection. As a result, the amount of spinning reserve increases with the growth of PV penetration. Battery storage can reduce the risk of PV's intermittent power supply, and always ensure demand satisfaction [2]. Earlier to the common use of alternate sources of energy, Economic Dispatch problem looked after only the conventional thermal power generators, which use non-renewable resources as fuels. Recently it has become necessary that there is a need for alternate method of generation apart from the conventional thermal energy power generation, and one of the sources that has gained popularity is the Solar-PV. Global trend of utilizing more renewable energy gave a path for the generation of electricity keeping in mind the environmental restrictions [4].

Solar Power generation model

The output power of photovoltaic (PV) is uncertain as it is mostly affected by the environmental factors, namely the environmental random changes will inevitably lead to constantly changing of output power of PV [25, 26]. Solar power depends on meteorological conditions such as irradiance, ambient temperature that are directly related to geographical location [8, 14]. For effective utilisation of PV arrays, the characteristics should be desperately analysed. For the proposed case study, Neural fuzzy will be used to predict the solar power output considering the insolation levels in a given day. A stochastic model [16] of Solar panel is constructed based on Beta distribution function. Beta distribution is considered to be the most suitable model for statistical representation of the probability density function.

III. ENERGY STORAGE SYSTEMS

Energy storage has the potential to provide a significant portion of the flexibility needed to manage the modern grid. These include supporting the overall reliability of the electricity grid, to help defer or avoid investments in other infrastructure, to provide backup energy during power outages or other energy shortages, to allow energy infrastructure to be more resilient, to support off-grid systems and to facilitate energy access for under-served populations. In 2016, a primary driver for advances in energy storage was the demand for battery storage in Electric Vehicles (EVs).

Energy storage systems (ESS) capture energy during periods when demand or costs are low, or when electricity supply exceeds demand, and can surrender stored energy when demand or energy costs are high [2]. Recent development and advances in the ESS and power electronic technologies have made the application of energy storage technologies a viable solution for modern power application [6]. The potential applications mainly cover the following aspects. Through time shifting, the power generation can be regulated to match the loads. The ESS can also be used to balance the entire grid through ancillary services, load following and load leveling [7]. Moreover, it can meet the increasing requirement of reserves to manage the uncertainty of wind generation [8] which can increase the system operation efficiency, enhance power absorption, achieve fuel cost savings and reduce Carbon emissions. Additionally, the ESS is a potential solution to smooth out the fluctuations, and improve supply continuity and power quality [9].

Energy storage used in conjunction with renewable energy resources for example wind and solar is one of the means to increase the use of renewable energy while maintaining a high quality of service reliability. The use of storage devices can help balance the wind and solar generation output and can be used to transfer energy from low-use periods to peak-use periods, allowing the system to operate at a more constant level and reducing energy supply costs [21]. Each technology has its own performance characteristics that makes it optimally suitable for certain grid services and less so for other grid applications. This ability of a storage system to match performance to different grid requirements also allows the same storage system to provide multiple services. This gives storage systems a greater degree of operational flexibility that cannot be matched by other grid resources, such as combustion turbines or a diesel generator.

A number of different energy storage technologies exist and are under development, and their characteristics (response time, discharge time, output capacity and efficiency) and functions vary widely. As of 2016, most electric energy storage capacity relied on pumped storage, the oldest and most mature electricity storage option, as well as the largest in scale (per system). Other electricity storage technologies include batteries (electro-chemical), flywheels and compressed air (both electromechanical). Thermal energy storage, which stocks heating or cooling for later use (e.g., molten salt, ice storage, etc.) also is present in some markets and can serve both thermal applications and electricity by conversion. Only pumped storage is a highly mature technology; all others are undergoing development and transition [2].

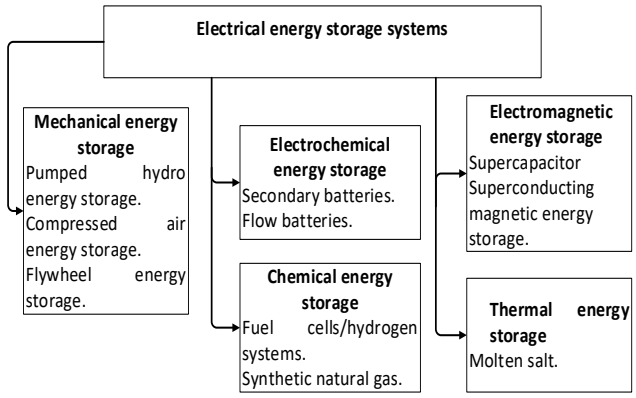


Fig. 1 Classification of electrical energy storage systems.

COMPARISON OF ENERGY STORAGE SYSTEMS

Pumped hydro storage and CAES have the lowest investment risks with respect to the cost per kilowatt-hour of electricity produced and the lowest levelised cost of delivered energy, comparable with combined cycle gas turbines. However, they are expensive to site and build, have long construction time, only suitable for large scales, have low energy density and typically result in long transmission distances. Lithium-ion batteries could potentially be a cost-effective option in the long term for short durations (less than 4 hours). Certain flow batteries such as Zn/Br and vanadium redox, and emerging options such as Fe/Cr and Zn/air, show potential for low cost in the 4 to 8 hour or longer energy duration range, however the cost of flow batteries could increase due to the associated cost of pumping the electrolyte. Fuel cells offer the highest energy density, however the capital cost is the largest among all energy storage options with one of the lowest energy conversion efficiency ranges [19].

In order to provide smooth and uninterrupted electricity supplies, it is necessary to combine different energy management strategies. Power quality management, load shifting and standby reserve are all-necessary to maximize the efficiency and reliability of the system. Each has very different desirable characteristics and the most appropriate energy storage systems will vary according to the desired role in the power management strategy. Power quality management relies on very fast response times to smooth electricity quality disturbances on a nanosecond and millisecond scale to provide uninterrupted, reliable power. The best examples for this application are flywheels, capacitors and SMES due to very fast response rates and the ability to be charged and discharged frequently whilst maintaining good operating lifetimes. Load shifting involves storing energy available in times of lower electricity demand and storing this for peak demand times. Thermal energy storage is well suited to load shifting due to low costs and good capacity, whilst batteries are most commonly used in this application. Standby reserve is an available reserve of power that can be brought online to take over from the main power-generating source if it should fail or become unavailable. Ideal energy storage systems must hold their charge for long periods and have the ability to operate for days without interruption. PHS and CAES have the largest capacity and low self-discharge, making them ideally suited to this application and the only technologies currently proven for utility standby reserve. Technologies under development that will likely be appropriate in the future include fuel cells and high temperature batteries such as Sodium Sulphur [23].

IV. ECONOMIC DISPATCH

The term "economic dispatch" refers to the practice of operating an electric system so that the lowest-cost generators are used first, followed by more expensive generators and then ramped down again when loads decrease. Economic dispatch must manage generation and demand resources efficiently over time. Electricity demand varies greatly, in daily, weekly and seasonal patterns. Because bulk electricity cannot be stored inexpensively at present, generation must be available to follow changes in load almost instantaneously, and some generation and demand reduction resources must be reserved to respond to sudden, unplanned contingencies, such as generator outages, as well as changes in customer demand and variable resource production levels [25].

Different generators have different costs, production capabilities and characteristics. A generator's production level at a point in time will be affected by how quickly it can safely move between output levels, whether it is operating in a high- or lower-fuel efficiency zone, fuel availability, and whether there is sufficient transmission capacity available to deliver its output across the grid. Grid operators adjust the output of dispatchable generators - including fossil, nuclear, geothermal and dam-impounded hydro frequently to reflect changing grid conditions. The costs associated with ramping large fossil generators up and down can be significant [26].

Increasingly, operators are looking to automatically dispatched demand-side resources and distributed storage devices to help manage small, short-term fluctuations in variable resource output. Variable generation resources like wind and solar photovoltaic are some of the fastest growing sources of capacity being integrated to the grid today. Generally, system operators accept as much electricity as possible from renewable resources, because of its low cost and only curtail reliance on these sources when forced to by limits on transmission availability or reliability considerations. While renewable generation adds variability and uncertainty to the system because the wind does not always blow, the sun does not always shine (variability), and we cannot perfectly predict when these changes will occur (uncertainty). Hence these units are not dispatchable in the traditional sense (i.e., cannot be precisely controlled by the grid operator), but their output is accepted as must-run or must-take production. [14]

V. PARTICLE SWARM OPTIMIZATION

Particle swarm optimization (PSO) also known as swarm intelligence is an evolutionary computation technique that has become a candidate for many optimization applications due to its high-performance and flexibility. Kennedy and Eberhart developed the PSO technique in 1995 based on the social behavior of flocking birds and schooling fish when searching

for food. This technique simulates the behavior of individuals in a group to maximize the species survival. Compared with other evolutionary algorithms, the main advantages of PSO are its robustness in controlling parameters and its high computational efficiency [12]

Each particle represents a member of the population and it "flies" in a direction that is based on its experience and that of the whole group. Individual particles move in the search space stochastically toward the position affected by the present velocity, previous best performance, and the best previous performance of the group. The neighborhood of any particle *i* is the subset of the particles it has capability to communicate with. Each particle has three parameters associated with it; its position in the search space, its velocity and the best position it has achieved individually. In every iteration, each particle keeps track of their personal best attempt, known as pbest and keeps track of their neighboring best and global best performance known as gbest. This knowledge is used further to know a better position (optimization of solution), this method combines self-experiences and social experiences. The main concept of PSO lies in the essence that each particle in the space accelerates towards the pbest and gbest locations with a random weighted acceleration in each iteration. Each particle updates its position and velocity according to the following equations [13].

$$v_i(t) = w_{ic}\ X\ v_i(t-1) + c_1\ X\ rand_1\ X\left(P_{best\ i} - x_i(t-1) + c_2\ X\ rand_2\ X\left(g_{best\ i} - x_i(t-1)\right)\right) \quad (1)$$

$$x_i(t) = x_i(t-1) + v_i(t) \quad (2)$$

Where;

t: is the current iteration

$v_i(t)$, $v_i(t-1)$ are current and previous iteration velocities respectively

$x_i(t)$, $x_i(t-1)$ are current and previous particle positions respectively.

c_1, c_2 are the learning factors.

$rand_1$, $rand_2$ are randomly generated numbers in the range [0.0,1.0]

$P_{best\ i}$: is the personal best position

$g_{best\ i}$: is the global best position

The inertia constant, w_{ic} is given as;

$$w_{ic} = w_{max} - \left(\frac{w_{start} - w_{stop}}{max_i}\right)\ X\ i \quad (3)$$

Where;

w_{max}: is the maximum inertia weight

w_{start} and w_{stop} are the inertia weights at beginning and end of an iteration respectively

max_i: is the maximum number of iterations

i: is the current iteration number

VI. OPTIMAL SIZING OF BATTERY ENERGY STORAGE SYSTEM

PSO will be used to optimally size the BESS and the following parameters will be selected for the proposed algorithm;

- The number of particles in the swarm NP
- The number of iterations Ni
- The learning factors c_1, c_2
- The maximum inertia weight w

The flow chart of the optimal BESS size based on Particle Swarm Optimization algorithm is shown in Fig. 2 below.

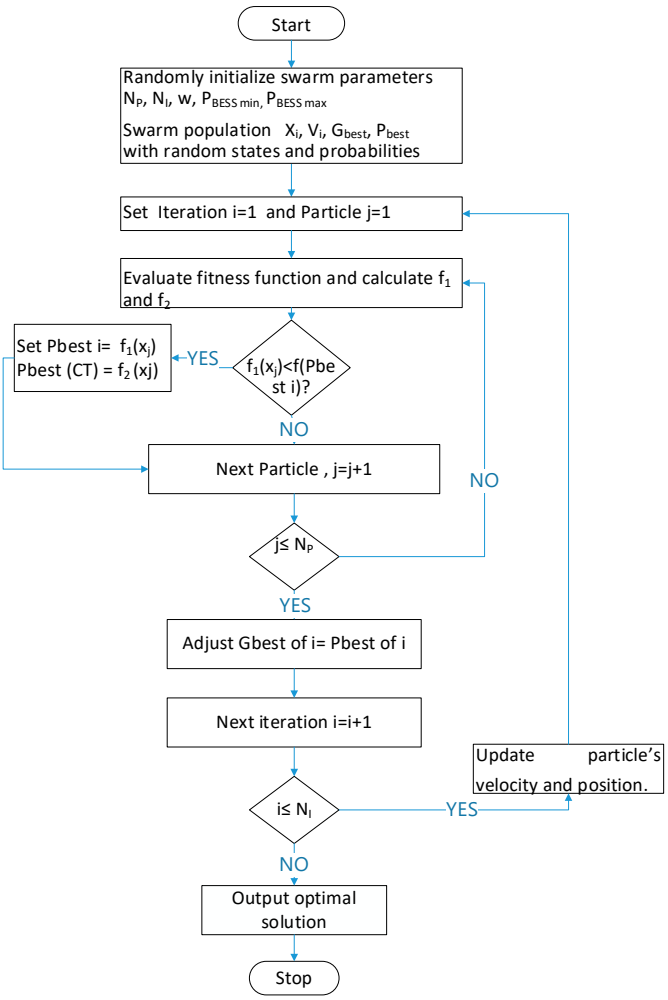


Fig.2 The overall flowchart of the proposed optimal BESS sizing based on PSO algorithm.

The BESS has to be sized optimally because an oversized BESS can be used to meet the desired features, which will lead to unnecessary high costs. On the other hand, an undersized battery may not meet the grid-connection standards. As a result, the right BESS design tool is important and beneficial for PV owners and power system planners

VII. PROBLEM FORMULATION

The PV-utility utilizes the energy generated by PV plants for minimizing the cost of operating thermal units. The PV system may be spread out over a large geographical area with a centralized location for battery. The short-term generation problem is to determine the hour for which thermal units of an electric power utility should be taken off-line or on-line. In order to produce a scheduling procedure that is practical it is

essential that numerous and complex constraints are incorporated into the solution method. The mathematical scheduling problem is formulated as follows,

$$Minimize\ TC = \sum_{i=1}^{N_t} F\big(P_T(t)\big) \quad (4)$$

Where

$F\big(P_T(t)\big)$: the cost of thermal unit *i* at time t

N_t: the number of thermal generation units

Subject to

$$P_T(t) + P_u(t) = P_l(t) \quad \text{for all t} \quad (5)$$

Where

$P_T(t)$: Power output of thermal unit at time t

$P_l(t)$: the load demand at time t

$P_u(t)$: Power output of the PV plant to the utility at time t

$$C(t) = C(t-1) + \left[\frac{\Delta t\,\eta_b(t)}{V_b(t)}\Big(P_{pv}(t) - P_u(t) - P_s(t)\Big)\right] \text{ for all t} \quad (6)$$

$$P_{pv}(t) - P_b(t) - P_u(t) - P_s(t) = 0 \text{ for all t} \quad (7)$$

$$P_{pv}(t) = f\big(G_l(t), T_C(t)\big) \text{ for all t} \quad (8)$$

Where

$C(t)$: current battery state of charge at stage t (Ah)

$C(t-1)$: previous battery state of charge (Ah)

$\eta_b(t)$: battery charge efficiency at stage t

$V_b(t)$: battery voltage at stage t (V)

$P_{pv}(t)$: PV unit power output at stage t (Mw)

$P_s(t)$: spillage power from PV at stage t (Mw)

$P_b(t)$: power charge/discharge to/ from battery at stage t (Mw)

$G_l(t)$: radiation on a tilted plane at stage t (W/m^2)

$T_C(t)$: PV cell temperature [^{o}K]

Penetration limit

$$P_u(t) < P_u^{max} \text{ for all t} \quad (9)$$

Charge/discharge limit

$$|P_b(t)| < P_b^{max} \text{ for all t} \quad (10)$$

Starting state of charge

$$C(t)|_{t=0} = C_s \quad (11)$$

Ending state of charge

$$C(t)|_{t=N_t} = C_f \quad (12)$$

State of charge limit

$$C_{min} < C(t) < C_n \text{ for all t} \quad (13)$$

Fixed charge for a particular hour

$$P_b(t)|_{t\in lm} = P_f \quad (14)$$

VIII. PROPOSED METHOD FOR ECONOMIC DISPATCH

The proposed system that has thermal units and photovoltaic (PV) unit with the battery system. PV generates the power proportional to the solar radiation; power from the PV unit is used to charge the battery during off-peak hours and it the battery is discharged during peak hours when demand is high. The spillage power is assumed to be zero. The output of the battery and the PV unit is DC hence the power is fed to the load via an inverter.

Initial Feasible Solution

The purpose of this algorithm is to determine the optimal battery size, which will produce optimal PV-battery generation in the initial schedule, and then the result will be used in economic dispatch problem. The objective function is to minimize the shortage of energy if there are no thermal units. The battery's State of charge (SOC) levels are used as state variables. Both starting and ending SOC, C, and C' are given along with loading periods. The constraints are available radiation and PV-battery limits.

$$Minimize\ ET = \sum_{t=1}^{N_t}\big(P_l(t) - P_u(t)\big) \quad (15)$$

Subject to;

$$C(t) = C(t-1) + \left[\frac{\Delta t\,\eta_b(t)}{V_b(t)}\Big(P_{pv}(t) - P_u(t) - P_s(t)\Big)\right] \text{ for all t} \quad (16)$$

$$P_{pv}(t) - P_b(t) - P_u(t) - P_s(t) = 0 \text{ for all t} \quad (17)$$

$$P_{pv}(t) = f\big(G_l(t), T_C(t)\big) \text{ for all t} \quad (18)$$

$$P_u(t) < P_u^{max} \text{ for all t} \quad (19)$$

$$|P_b(t)| < P_b^{max} \text{ for all t} \quad (20)$$

$$C(t)|_{t=0} = C_s \quad (21)$$

$$C(t)|_{t=N_t} = C_f \quad (22)$$

$$C_{min} < C(t) < C_n \text{ for all t} \quad (23)$$

$$P_b(t)|_{t\in lm} = P_f \quad (24)$$

Economic Dispatch problem

The main objective of economic load dispatch problem is to minimise the fuel cost of the thermal generators [9]. The objective function can be formulated as;

$$F_i(P_i) = \sum_{i=1}^{N_t}(a_i P_i^2 + b_i P_i + c_i) \ \$/h \quad (25)$$

Where

$F_i(P_i)$: Fuel cost of all the generators ($/h)

a_i, b_i, c_i: Cost coefficients of ith generator.

P_i : Power generated by ith generator (Mw)

Suppose a PV-utility plant has N thermal units, the total load of thermal units at each stage is;

$$P_T(t) = P_l(t) - P_u(t) \quad (26)$$

The constraints of the optimization problem are;

Equality constraints:

The power balance of the system is given by;

$$\sum_{i=1}^{N_t} P_i(t) - P_D(t) - P_L(t) = 0 \quad (27)$$

where

P_D : Load demand (MW)

P_L: Transmission losses (MW)

Transmission losses can be represented as;

$$P_L(t) = \sum_{i=1}^{n}\sum_{j=1}^{n} P_i(t) B_{ij} P_j(t) \quad (28)$$

where

B_{ij}: Transmission loss coefficient

Inequality constraints:

The power generation of all the generators has maximum and minimum limits

$$P_i^{min} \le P_i(t) \le P_i^{max} \quad (29)$$

Where

P_i^{min} and P_i^{max} are the minimum and maximum generation limits respectively.

Thermal unit minimum starting up/down times

$$\big(X_t^{on}(t-1) - T_i^{on}\big)\big(I_i(t-1) - I_i(t)\big) \ge 0 \quad (30)$$

$$\big(X_t^{off}(t-1) - T_i^{off}\big)\big(I_i(t-1) - I_i(t)\big) \ge 0 \quad (31)$$

Ramp rate limits

$P_i(t) - P_i(t-1) \leq UR_i$ as unit i ramps up (32)

$P_i(t-1) - P_i(t) \leq DR_i$ as unit i ramps down (33)

Dynamic Economic Dispatch

The objective of the dispatch problem is to achieve the minimum production cost by properly scheduling of PV units, thermal units, battery energy storage units for the given period. The results must meet system constraints. Given the available thermal units in each stage, we solve the PV-thermal scheduling problem. The output power from a solar panel depends mainly on the irradiance. So, the power output for various irradiance values is to be estimated which requires proper functional model. The best adopted model is beta distribution function. The historical data of solar irradiance is processed and then it is utilised for modelling the beta distribution function. Using this function the output of a solar panel is estimated and then the total output obtained for the entire solar farm is calculated. This power generated by the solar farm is considered as negative demand and is incorporated at the specific point.

Then economic dispatch is carried out using this model and the results are compared. For incorporating the solar energy into the exiting generation, the power generated by PV arrays is considered as a negative load and the load equation is updated as follows:

$P'_D = P_D - \sum_{iS=1}^{n} P_{iS}$ (34)

where

P'_D : new power demand (MW)

$\sum_{iS=1}^{n} P_{iS}$: the sum of solar power generators (MW)

IX. Conclusion

The paper presents a sizing methodology to find the optimal battery energy storage capacity and power Particle Swarm Optimization algorithm to minimize the cost of investment and losses incurred by the system in form of peak load shaving. Then Economic Dispatch for a utility incorporated with PV-battery be conducted, and by incorporating battery storage, we can reduce load following requirements in the PV-utility grid. Furthermore, peaking generators can be kept off during peak hours by utilizing PV-battery. The formulation developed in this paper is very flexible and can be applied to other renewable energy sources with intermittent nature. Thus BESS makes PV power dispatchable in a similar manner to conventional sources. This show a pathway for the use of renewable energy in future for solving Economic Dispatch problem more effectively.

Appendix

Power generation from PV array

The expected output of PV is given by;

$P(S) = P_o(S) * f_b(S)$ (35)

The total output of the PV array corresponding to specific time segment is given by;

$TP = \int_0^1 P_o(S) * f_b(S) dS$ (36)

where power generation of panel at solar irradiance s is given by

$P_o(S) = N * FF * V_y * I_y$ (37)

where

N: the total number of PV modules.

The voltage - current characteristics of a PV module for a given radiation level and ambient temperature are determined using the following relations

$FF = \frac{V_{MPPT} * I_{MPPT}}{V_{OC} * I_{SC}}$ (38)

$V_y = V_{OC} - K_v * T_{cy}$ (39)

$I_y = S\left[I_{SC} + K_i\left(T_{cy} - 25\right)\right]$ (40)

$T_{cy} = T_A + S\left(\frac{N_{OT}-20}{0.8}\right)$ (41)

where

FF: the fill factor,

V_{MPPT}: the voltage maximum power point

I_{MPPT}: the current maximum power point,

V_{OC}: the open circuit voltage

I_{SC}: the short circuit current of PV module

K_v: the voltage temperature coefficient and

K_i: the current temperature coefficient,

T_A : the ambient temperature

T_{cy}: PV cell temperature

N_{OT} : normal operating temperature

References

[1] S. Abedi, G. H. Riahy, S.H.Hosseinian, and A. Alimardani, "Risk-constrained unit commitment of power system incorporating pv and wind farms"

[2] R. E. P. Framework, "Renewables global status report: 2009 update," Renewable Energy World, 2009. View at Google Scholar.

[3] M. Shahidehpour and F. Schwarts, "Don't let the sun go down on PV [photovoltaic systems]," Power and Energy Magazine, vol. 2, pp. 40–48, 2004. View at Google Scholar

[4] M.Narender, J.Bhagawan and M.Mustafa, "Economic load dispatch of thermal and pv system with battery storage"

[5] T. O. Ting, M. V. C. Rao, and C. K. Loo, "A novel approach for unit commitment problem via an effective hybrid particle swarm optimization," IEEE Transactions on Power Systems, vol. 21, no. 1, pp. 411–418, 2006.

[6] M. Shahidehpour, H. Yamin, and Z. Li, "Market operations in electric power systems," Wiley, New York, NY, USA, 2002.

[7] V. P. Pappala, I. Erlich, K. Rohrig, and J. Dobschinski, "A stochastic model for the optimal operation of a wind-thermal power system," IEEE Transactions on Power Systems, vol. 24, no. 2, pp. 940–950, 2009.View at Publisher · View at Google Scholar · View at Scopus

[8] Y. Yang "Optimization of battery energy storage systems for pv grid integration based on sizing strategy"

[9] Banerjee, Sumit, D.Maity, and C.K.Chanda. "Teaching learning based optimization for economic load dispatch problem considering valve point loading effect." International Journal of Electrical Power & Energy Systems 73 (2015): 456-464.

[10] Y.M. Atwa, et al. "Optimal renewable resources mix for distribution system energy loss minimization." Power Systems, IEEE Transactions on 25.1 (2010): 360-370.

[11] Kayal, Partha, and C. K. Chanda. "Optimal mix of solar and wind distributed generations considering performance improvement of electrical distribution network." Renewable Energy 75 (2015): 173-186.

[12] P.J. Angeline, "Evolutionary optimization versus particle swarm optimization: philosophy and performance differences", Lecture Notes in Computer Science, 1998, Vol. 1447, pp. 601–610

[13] D.W. Boeringer and D.H Werner "Particle swarm optimization versus genetic algorithms for phased array synthesis", IEEE Transactions on Antennas Propagation, 2004, Vol. 52, pp. 771-779

[14] 2011/2012 Economic dispatch and technological change report to congress September 2012; US department of energy Washington DC 20585.

[15] J. Kondoh, I.Ishii, H. Yamaguchi, et al. “Electrical energy storage systems for energy network.” Energy Conservation Management 2000; 41(17):863–1874.
[16] IEC. Grid integration of large-capacity renewable energy sources and use of large-capacity electrical energy storage, Tech report; 2012.
[17] T. Boutsika and S. Santoso. “Sizing an energy storage system to minimize wind power imbalances from the hourly average.” In: IEEE PES general meeting, San Diego, California, USA, July 22–26, 2012.
[18] R.Billinton, R.Karki, Y.Gao, et al. “Adequacy assessment considerations in wind integrated power systems”. IEEE Trans Power Syst 2012; 27(4):2297–305.
[19] IEC. Electrical energy storage white paper. Tech report; 2011.
[20] Chen H, Cong TN, Yang W, et al. “Progress in electrical energy storage system: a critical review.” Progr Nat Sci 2009; 19 (3):291–312.
[21] [13] M. Swierczynski, R. Teodorescu R, Rasmussen et al. “Overview of the energy storage systems for wind power integration enhancement. In: Proceedings of IEEE international symposium on industrial electronics, Bari, Italy, July 4–7, 2010.
[22] “Energy storage for grid connected wind generation applications.” EPRI-DOE handbook supplement; 2004.
[23] D. Connolly, “An Investigation into the energy storage technologies available for the integration of alternative generation techniques.” Tech report; 2007.
[24] J.A. Wood and F. B. Wollenberg “Power Generation, Operation & Control” 2nd edition, Wiley India, New Delhi, 2010.
[25] J.J.Grainger and W.D.Stevenson Jr, “Power system analysis,” Tata McGraw Hill, New Delhi, 2012

Machine to Machine (M2M) Communications for Dynamic Pricing in Smart Grid Networks

Kenneth Kimani[1*], Kibet Langat[2] and Vitalice Oduol[3]

***Abstract*—**Smart grid (SG) networks are characterized by a tight integration of a flexible and secure communications network supporting a very large number of sensor and actuator nodes. This communication network is vital in learning about the power consumption patterns of consumers and finding out the statistical characteristics of power sources. With the accurate information, the smart grid can plan, support and interact with consumers in a more efficient and economical way. In order to realize the intelligent electricity network, machine-to-machine (M2M) communication is considered as a core building block for the SG enabling the deployment of a wide-scale monitoring and control infrastructure. The integration of machine to machine communications in the grid network eliminates human related errors by automating the processes of reading and reloading of utility meters, collecting information on customer usage patterns and instantaneously informing the utility companies in case of any faults and/or outages in the network.

M2M communications enables new applications in the smart grid including smart metering and dynamic pricing of power in the grid. Smart metering in M2M facilitates flexible demand management where a smart meter measures consumption and communicates that information back to the local utility. Dynamic pricing is used to control power demand in response to fluctuation in power supply. When the power demand from customers is higher than the power production output, the utility provider needs to source for additional power supply from other sources such as diesel operated power plants which are more expensive and harmful to the environment, whereas, when the power demand is too low, the operator may suffer from an under-utilized infrastructure leading to wasted capacity.

M2M communications enables the effecting of dynamic pricing structures whereby customers are charged higher rates during peak hours to discourage much consumption and charged lower rates during off-peak hours to encourage consumption. The different levels of tariffs allow the consumer to make a smarter and more responsible choice and can be used as one of the mechanisms of effecting load balancing which is a very big hurdle in existing grid networks. In this paper, we discuss the application of M2M communications to the SG to achieve dynamic pricing structures. We also explore the challenges and opportunities achieved from the dynamic pricing in the smart grid.

***Keywords*—** Advanced Metering Infrastructure, Dynamic Pricing Structures, Machine-to-Machine (M2M) Communications, Smart Grid, Smart Meters.

I. Introduction

Current smart grids are majorly composed of machine-to-human or human-to-human information production, exchange, and processing. However, due to the large geographical spread of grid networks coupled with the large numbers of devices and the resulting large amounts of data, this has proved to be unsustainable resulting in inefficiencies such as in meter transactions (meter reading in analog postpaid meters). Also, in most places especially in the rural areas, there still exist analog utility meters where the utility companies have to send manpower to go and physically collect the meters' readings. Even in places where the digital prepaid meters are said to exist, they still require human input to manually key in the prepaid tokens into the meter.

Recent years have seen an upsurge in the number of new electricity connections and as more and more energy consumers are connected to the power grid, energy demand is expected to be more volatile due to fluctuation of the demand at the power network, combined with potential variations in generation and the unpredictability of the availability of distributed renewable sources [1]. This necessitates a paradigm shift on how we collect, process and control data in the smart grid considering the large number of resources in a smart grid environment and the size of the data to be moved.

A significant amount of intelligence needs to be injected in the current grid networks through the incorporation of information and communication technologies to dramatically improve (or even change) the management and operation of every aspect of the power system. In order to realize the intelligent power network, machine-to-machine (M2M) communication is the most ideal to deploy a wide-scale monitoring and control infrastructure in order to enable real-time interaction with the power grid network [2].

The integration of machine to machine communications in the grid network is envisioned to eliminate human related problems by automating the various processes in the grid network [3]. The implementation of the M2M communication system in the smart grid network also paves the way for effecting of dynamic pricing structures whereby customers are charged higher rates during peak hours to discourage much consumption (consumption by critical uses) and charge lower rates during off-peak hours to encourage consumption which can then be deployed as a load balancing mechanism.

Machine to Machine (M2M) communication allows communication to take place amongst machines such as sensors, smart meters, and other equipments in the smart grid [4] enabling them to maintain a near real-time awareness of

Kenneth Kimani, Department of Telecommunications & Information Engineering, JKUAT (phone: +254724012663; e-mail: kuriakimani@yahoo.com).

Kibet Langat, Department of Telecommunications & Information Engineering, JKUAT (e-mail: kibetlp@jkuat.ac.ke).

Vitalice Oduol, Department of Electrical & Information Engineering, UoN (e-mail: vkoduol@uonbi.ac.ke).

each other's operating requirements and capabilities without requiring human intervention. It is estimated that the amount of generated energy-related data will be up to tens of thousands of terabytes in the near future [5]. This poses a significant challenge for any existing communication network as well as future smart grid networks [6]. M2M communication supports large scale data transmissions and is envisaged as a solution where the power consumption of different domestic devices can be monitored and various grid services such as real time power pricing, power consumption scheduling and generation can be managed and controlled in real time by using the data collected from the smart meters [7]. The presence of the machine to machine communications network provides energy efficiency in the smart grid by enabling the balancing of the generation and consumption levels, thereby streamlining the production and consumption of power in the grid [8] .

The rest of the paper is organized as follows. In Section II, Machine to Machine Communications is introduced. The effect of load imbalance on the grid network is discussed in section III. Section IV discusses various methods of performing dynamic pricing in the smart grid. Section V explores the network architecture of the smart grid. The requirements for M2M communication in the smart grid are explored in section VI. In Section VII, the challenges of the smart grid that are solved with the adoption of M2M communications in the smart grid are summarized. Finally, the concluding remarks are presented in Section VIII.

II. MACHINE-TO-MACHINE (M2M) COMMUNICATIONS

Machine-to-machine (M2M) communications is an emerging communication paradigm that provides ubiquitous connectivity between devices along with an ability to communicate autonomously while requiring no human intervention [9]. M2M communications are characterized by fully automatic data generation, exchange, processing and actuation among intelligent machines, without or with low intervention of humans [4] [10]. The objective of M2M communications is to increase the level of system automation in which the devices and systems can be interconnected, networked, and controlled remotely, with low-cost, scalable, and reliable technologies in order to exchange and share data [11].

The European Telecommunications Standards Institute (ETSI) has proposed an end-to-end architecture for M2M communications [12] comprised of five key elements:

- **i)** The M2M component, which is usually embedded in a smart electrical device, replies to requests or transmits data.
- **ii)** The M2M gateway enables connectivity between the M2M components and the communication network.
- **iii)** The M2M server works as a middleware layer to pass data through various application services.
- **iv)** The M2M area network provides connectivity between M2M components and M2M gateways.
- **v)** The M2M communication network provides connection between M2M gateways and is M2M servers.

These five elements constitute the M2M system as specified by ETSI: the M2M component working in the device domain, the M2M area network and gateway in the network domain, and the M2M server and communication network in the application domain as shown in figure1 below. From the data management perspective, M2M communications consists of three phases: data collection, data transmission, and data processing. The data collection phase refers to the procedure used to obtain the physical data. The data transmission phase includes the mechanisms to deliver the collected data from the communications area to an external server. The data processing phase is the process of dealing with and analyzing the data, and also provides feedback on how to control the application.

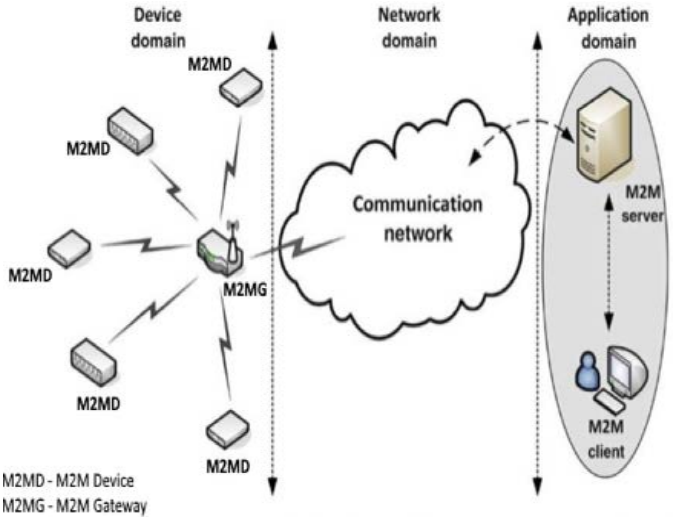


Figure 1: Architecture of M2M Networks [9]

In the context of M2M communications, the M2M server is located at the control center. The smart meters form the bulk of the M2M devices in a SG network. A smart metering service is considered to be one of the most prominent M2M applications, and is an advanced way to measure the consumption of electricity [4] [9]. A large number of metering machines monitor and record the energy consumption and automatically report to the smart metering server through M2M communications [13] [14].

The semiconductor industry's shrinking lithography and improving yield have led to reduced chipset cost and low power consumption leading to a rise in the number of connected devices that are not mobile phones and do not require human control and consequently leading to an increase in number of

M2M devices. Some of the most prominent M2M application areas include smart grids (grid control, industrial metering, demand response), vehicular telematics (fleet management, enhanced navigation, etc.), healthcare (telemedicine, remote diagnosis, etc.), manufacturing (production chain monitoring), and remote maintenance (industrial automation, vending machine control, etc.) [9].

III. Load Balancing

Load balancing is a process which involves resource management and an effective load distribution among the resources to ensure maximum utilization of resources is achieved. Kenya Power (Kenya's main power distribution utility company) has an estimated five million customers as of December 2016. Based on statistics from KPLC, the peak electrical load was 1228 MW and the total yearly electrical demand was 7.53 TWh, corresponding to an average demand of 860 MW. This shows that load balancing is a big issue for the utility provider. Figures 2 and 3 below show the average daily load profile and the average weekly load profile respectively.

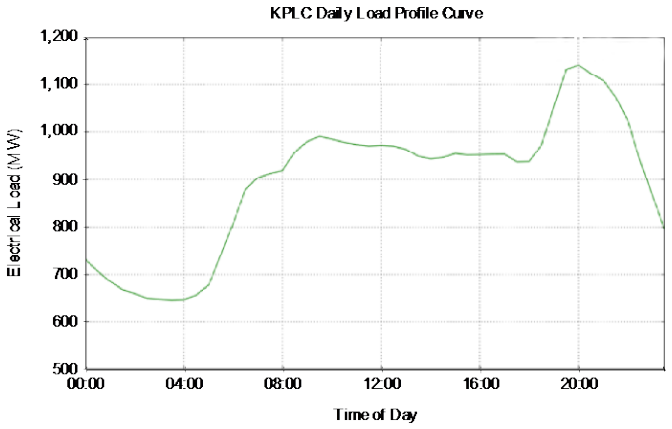


Figure 2: Average daily load profile curve

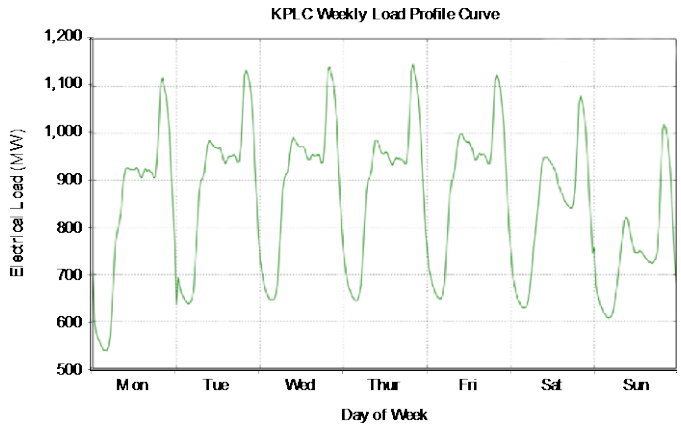


Figure 3: Average weekly load profile curve

The load profile curves show the peak time to be from 6 pm to 10 pm with the highest peak at 1140 MW at 8 pm compared to the off-peak period of between 12 am to 6 am with the lowest at 650 MW at 3am. If approximated demand is higher than actual demand, the supplied power is wasted (i.e., over-supply). On the other hand, if actual demand is higher than the approximated demand, additional power supply is required (i.e., under- supply).

Due to intermittent power supply from some renewable electricity sources such as wind power, tidal power, solar power and in absence of large scale, grid wide energy storage technology, there are periods in the load profile when there is excess demand beyond the supplied production levels resulting in load imbalance. Most of the current load balancing mechanisms concentrate on the use of peaking power plants to fill in demand gaps [15]. Theses peaking plants mostly run on diesel and are run only when there is a high demand for electricity. These diesel power plants are known to give out emissions that cause environmental pollution. They also generate power at a much higher price per kilowatt hour than base load power resulting in higher price of electricity. Load Factor is a measure of the efficiency of electrical energy usage; a low load factor is an indication that load is not putting a strain on the electric system, whereas a high load factor is an indication that the load is putting a strain on the electric system.

The Load Factor f_{LOAD} is given by

$$f_{LOAD} = \frac{Total\ Load}{Maximum\ Load\ in\ a\ given\ Time\ Period} \quad (1)$$

Calculating the load factor for Kenya power, we find that

Peak Demand = 1228 MW

Total Yearly Electrical Load = 7.53 TWh

Number of days in billing cycle = 365

Load Factor (%) = {7530000 MWh / (365 days × 24 hours per day × 1228 MW)} × 100% = 69.99%

The process of load balancing in grids can be generalized into the following four basic steps as shown in figure 4 below.

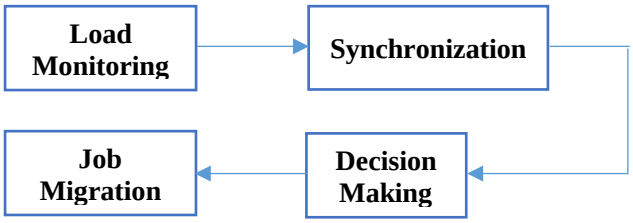


Figure 4: load balancing steps.

(i) Load Monitoring: Monitoring the resource load and state.
(ii) Synchronization: Exchanging load and state information between resources.
(iii) Decision Making: Calculating the new work distribution and making the work moment decision.
(iv) Job Migration: Actual data movement.

These four steps can easily be automated and synchronized

by applying an M2M communication network.

IV. Dynamic Pricing in the Smart Grid

Energy pricing is based on tariffs. In the Kenyan context these tariffs are set by the Energy Regulatory Commission (ERC). In this case, energy tariffs are set based on the Fuel Cost Charge(FCC), the Foreign Exchange Rate Fluctuation Adjustment (FERFA), the Inflation Adjustment (IA), the Water And Resources Management Authority (WARMA) levy, the Energy Regulatory Commission levy (ERC) and the REP levy. This approach is largely a static pricing structure where the prices are adjusted once a month. However, this approach fails to take into account the instability on the grid network brought about due to load shifting during peak hours and off-peak hours [16].

Electricity pricing policies can be either static or dynamic. Static prices do not vary with a variation in demand, whereas dynamic prices change with changing demand levels. The main pricing policies are identified as:

a. **Flat tariffs**: In flat-tariff power pricing structures, all the consumers are charged the same price per unit of power throughout the day. The price therefore remains static even though power demand varies over the period of consideration. Consumers under this pricing scheme do not suffer from changing costs of power supply with a change in aggregate demand levels. The consumers do not face any risk of high electricity bills for any unavoidable or unplanned electricity consumption [16]. Therefore, they have no financial motivation to change their energy usage. Due to its static nature, this scheme is often viewed as a welfare pricing scheme.

b. **Block Rate tariffs**: This pricing scheme differentiates between customers based on the different blocks of quantified electricity consumption. In this scheme, the electricity consumption is grouped into multiple tiers based on the amount of consumption. With each increasing block of consumption the per-unit rate increases correspondingly.

c. **Seasonal tariffs**: This pricing scheme is based on charging different rates in different seasons to match the varying demand levels between different seasons. Electricity is charged at a higher rate during high-demand seasons and a lower rate during low-demand seasons.

d. **Time-of-use (TOU) tariff**: This is a tariff based on pre-declared tariffs that vary during different times of the day. This tariff is also known as time-of-day (TOD) tariff. The TOD tariffs are usually high during peak hours and low during off-peak hours and can be effected for either short or long terms [16].

e. **Superpeak TOU**: This tariff is similar to TOU, but the peak window is shorter in duration (about four hours) so as to give a stronger price signal.

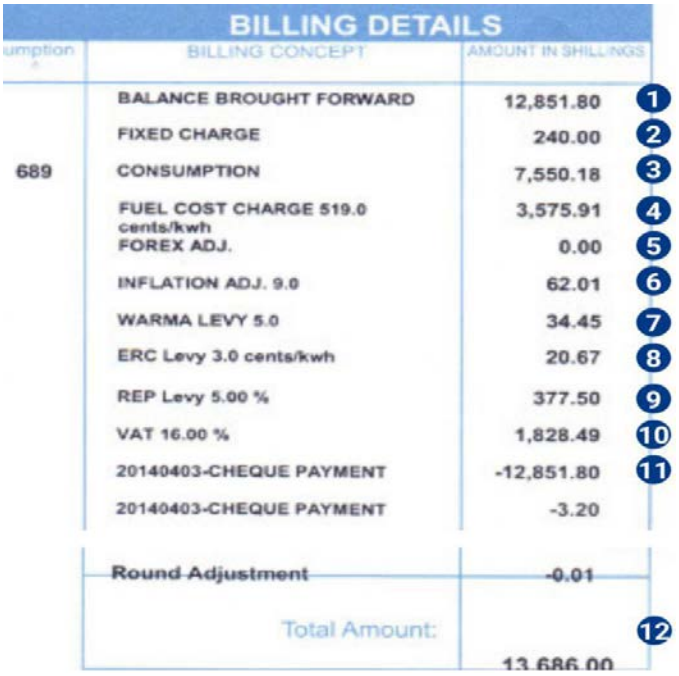
f. **Critical peak pricing (CPP)**: This is a dynamic pricing scheme in which prices are high during a few peak hours of the day and discounted during the off-peak hours of the day. This scheme helps in reduction of excessive peak load and its peak price remains relatively uniform for all days.

g. **Variable peak pricing (VPP)**: This tariff is similar to CPP but the only difference is that the day to day peak prices vary per day. However, the consumers have to be informed about the peak prices beforehand [16].

h. **Real-time pricing (RTP)**: This tariff is the purest form of dynamic pricing. In this tariff, prices change at very small intervals in time thereby posing the greatest uncertainty to the consumers. It increases the efficiency of dynamic pricing by reflecting the actual supply cost and actual level of demand at any particular instance. However, in order to manage these frequent changes, it requires an advanced communication technology due to the high rate of data collection and transfer.

i. **Peak time rebates (PTR)**: In this tariff rebates or discounts are provided to consumers for consuming electricity below a certain predetermined level during peak hours. These rebates can then be redeemed at a later time.

Figure 5 below shows a sample of a bill from Kenya Power showing the pricing structure employed in calculating the energy bill.



	BILLING CONCEPT	AMOUNT IN SHILLINGS	
	BALANCE BROUGHT FORWARD	12,851.80	1
	FIXED CHARGE	240.00	2
689	CONSUMPTION	7,550.18	3
	FUEL COST CHARGE 519.0 cents/kwh	3,575.91	4
	FOREX ADJ.	0.00	5
	INFLATION ADJ. 9.0	62.01	6
	WARMA LEVY 5.0	34.45	7
	ERC Levy 3.0 cents/kwh	20.67	8
	REP Levy 5.00 %	377.50	9
	VAT 16.00 %	1,828.49	10
	20140403-CHEQUE PAYMENT	-12,851.80	11
	20140403-CHEQUE PAYMENT	-3.20	
	Round Adjustment	-0.01	
	Total Amount:	13,686.00	12

Figure 5: Current Billing structure from Kenya Power

Dynamic pricing is where consumers are charged varying prices depending upon the demand response curve. Dynamic pricing induces consumers to adjust their demand in response to the time-varying prices by switching off some of the non-urgent, non-critical loads. For example, dish washers in some households have a delayed start option. And yet, only a small

fraction of consumers utilizes this feature and use the appliance later at night at off-peak times. This is bound to become more attractive with the rapid adoption of electric cars as more people are bound to be more interested in charging their cars during cheaper times of the night.

Demand response in a smart grid is expected to offer economic benefits to consumers while improving overall operation efficiency and reliability. A properly designed demand response program can reduce the peak load, compensate for uncertainties associated intermittent renewables, and reduce the cost of system operation. Demand response can easily be effected through through dynamic pricing mechanisms. Figure 6 below shows how the market forces of demand and supply are left to determine the equilibrium price at any particular point in time in case of RTP.

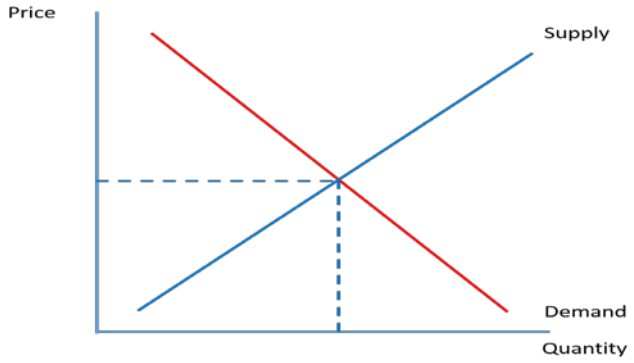


Figure 6: Demand - Supply price equilibrium curves

However, in order to implement this kind of time-sensitive pricing, utilities have to deploy more advanced smart meters to track how much energy people are using during certain times. This is a major reason why dynamic pricing is not more common. The availability of real-time pricing gives both consumers and suppliers' valuable indications to help manage their energy demands and supplies, respectively. Pricing applications broadcast pricing information to smart meters and smart appliances.

Other than demand-supply curve, a crucial factor that is considered in effecting dynamic pricing is market segmentation. Segmentation of the electricity market helps in differentiating customers based on various attributes. Attributes of market segments are helpful in setting the range of prices or the time span for maintaining a certain price in a dynamic pricing environment. Segmentation can be based on various demographic, behavioral and geographic factors. Current Segments are based on consumption data - high level and low level. There is also segmentation according to the market utilization – industrial and residential segments.

Important considerations that warrant for segmentation of electricity markets include welfare of vulnerable groups in society as these groups would be worst affected in case of improper dynamic pricing implementation. The effect of dynamic pricing schemes on low-income consumers show that low-income consumers are more sensitive to price signals than high-income ones. However, they are noted to have the lowest price responsiveness. This is because they have fewer opportunities to reduce consumption due to unavailability of specific home appliances in which the energy consumption can be controlled.

V. Communication Network Architecture for the Smart Grid

The SG has 4 main functional domains including power generation, transmission, distribution and power consumption [17]. However, in terms of the communication architecture, the SG is organized into three three-layered hierarchical domains consisting of Smart Grid Home Area Networks (SG-HANs), Smart Grid Neighborhood Area Networks (SG-NANs) and Smart Grid Wide Area Networks (SG-WANs) [9] as shown in figure 7 below.

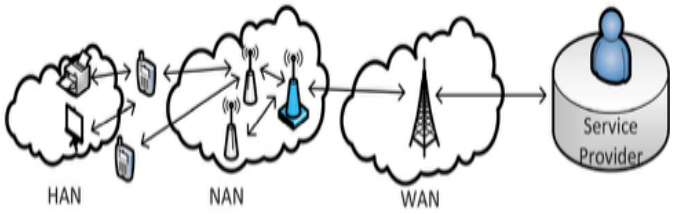


Figure 7: Hierarchical structure of SG communications [18]

a) Smart Grid Wide Area Networks (SG-WANs)

This is the upper layer of the SG communication architecture and serves as the backbone for communication between network gateways or aggregation points, NANs, SG substations, distributed grid devices, and the and the core utility systems. Both wireless and wired communication technologies can be used for communication requirements of the grid elements in SG-WAN [9].

b) Smart Grid Home Area Networks (SG-HANs)

This is the bottom layer of the hierarchical architecture within the SG dedicated to effectively managing the on-demand power requirements and consumption levels of the end-users. It is located in the customer domain and enables home automation networks for monitoring and control applications for efficient home energy management and user comfort. It consists of smart meters, sensors, actuators, and other intelligent devices around the home environment [5].

c) Smart Grid Neighborhood Area Networks (SG-NANs)

This is the intermediate layer of the SG communications

architecture that connects multiple SG-HANs to the SG-WAN of the utility company [5]. A smart grid neighborhood area network communicates and manages several SG-HANs within its coverage area. NANs gather the energy consumption information from multiple HANs and transmits the information to utility center through the upper layer of the communication architecture, i.e. the SG-WAN [9].

These smart grid communication network domains can be implemented using a variety of communication technologies with M2M communication providing the base of any network layout implemented in the smart grid.

VI. Requirements of the M2M communication in SG.

M2M communications plays an important role in data exchange and can be adopted in many applications with the sole objectives of improving efficiency and reducing cost. The smart grid is one of the strongest driving forces for the advance of M2M communications [19]. The M2M communication architecture for the smart grid must therefore meet the needs of current and future applications in a cost-effective, scalable, reliable, and secure way. Several key features make M2M communications to be ideal for deployment in smart grid networks and are needed to efficiently support M2M traffic characteristics [20].

1. Support for mass device transmissions

This feature allows for the handling of simultaneous transmission attempts from an extremely large number of M2M devices so that the network is not overwhelmed by the large amount of devices attempting to communicate. The support for multiple device transmissions as offered by M2M communications is very attractive for smart grids.

2. Extremely low power consumption

Some devices on the smart grid are located in locations with limited supply of power while others are battery operated. Some devices especially in power generation are located in remote areas that experience infrequent human interaction. Due to the large number of devices in the smart grid it would be very uneconomical to have to replace the battery every so often considering that the grid covers large geographical regions. Most M2M devices are low power devices hence increasing the life span of these devices on the grid.

3. Small bursts data transmissions

Support for transmission of small bursts of information by M2M communications resonates with the type of traffic found in smart grids. Smart grid traffic consists of small bursts transmissions of mostly control and monitoring data.

4. Group control

Group control implies that the system supports group addressing and handling of M2M devices. Enabling group control of mass devices in smart grids is very important as it allows the easy broadcast of changes in dynamics in the grid network. Such changes include changes in pricing information due to uneven supply and demand which might lead to load imbalance in the grid network. Group control allows for easy addressing of extremely large number of devices**.**

5. Interoperability

A key feature of smart grids is the interconnection of a potentially large number of disparate energy distribution networks, power generating sources and energy consumers. The components of each of these entities will need a way of communicating that will be independent of the physical medium used and also independent of manufacturers and type of devices. As a result, multiple communication technologies and standards could coexist in different parts of the system. Interoperability will help to cater for the highly heterogeneous traffic patterns of most smart grid systems.

6. Scalability

With billions of embedded devices expected to be integrated in the smart grid in the next couple of years, it is important for the network to have capacity and spectrum efficiency to support a large increase in number of connected devices while still maintaining acceptable Quality of Service (QoS) standards. Such a steep expansion in the M2M connected devices presents a great challenge for the network infrastructure that must now be able to cope with an abundance of devices [21].

7. Delay-tolerant traffic

Time-tolerant traffic can support significant delays in data transmission and reception. This implies that the system can give lower access priority to or defer data transmission of time-tolerant traffic. This feature enables simplifications to the bandwidth request.

8. Extremely low latency

Extremely low latency requires that both network access latency and data transmission latency be reduced. This feature is required in many emergency situations. Changes to the bandwidth request and network entry/re-entry protocols may be required to support extremely low latency.

9. High reliability

High reliability implies that connectivity and reliable transmission should be guaranteed regardless of operating environment (e.g., mobility, channel quality). This is vital in supporting instant billing mechanisms where payments made should should be able to reflect automatically. Improved reliability may require changes to the link adaptation protocol or modulation/coding schemes. Other solutions may involve improved interference mitigation, device collaboration, or redundant path establishment [22]. M2M devices are also expected to be low cost, so that they can easily be embedded in real fields and extensively deployed in a large scale.

10. Priority access mechanism

Priority access is necessary in order define a scheme of prioritizing access to resources in the smart grid communication network. This is vital in communicating monitoring information that might pose a danger to the power consumers. It also avoids losses especially in an era of limited resources such as bandwidth [23].

VII. Mapping of M2M domains onto the Smart Grid

M2M communication allows data transfer for direct machine-to-machine communication with little or no human interaction [24]. Whereas current grid systems are optimized more for human-to-human (H2H) communications, M2M communications is designed for low power consumption, large number of devices, long-range and short-burst data transmissions. This informs the need to switch from current analog networks to M2M communications [13] [14], in which devices build a virtual mesh that is connected to a central aggregator. M2M communication for smart grid applications provides a communication architecture that is much better than current smart metering technologies as shown in figure 8 below.

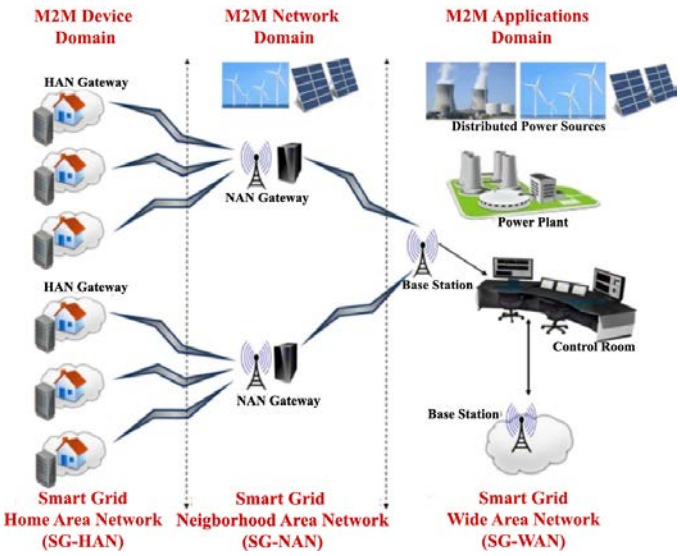


Figure 8: Mapping of the smart grid network onto the M2M architecture [25].

M2M communications architecture is hierarchical in that it comprises specific network segments for households, buildings or group of households, and neighborhoods, aiming at being flexible, adaptive, and scalable [11]. This hierarchical heterogeneous communications architectures comprising several network segments and combining different communications technologies present higher flexibility, so they fit a wider range of Smart Grid applications with their specific requirements and constraints [19]. This is the approach followed to design the M2M communications architecture and which is also aligned with the solution proposed as shown in figure 8 above.

The M2M communication network for the smart grid must be designed considering the scalability and energy efficiency requirements, periodic traffic patterns, and large-scale deployments. Optimal network design while minimizing the cost of M2M communications and data processing due to large amount of information collected is one of the main challenges hindering the full integration of M2M devices in the smart grid [26].

VIII. Conclusion

As the smart power grids continue to evolve, it is necessary to consider the most robust and reliable technology to facilitate communication in the grid networks. M2M communications play an important role in data exchange in the smart grid thereby enabling new services and applications. One of these applications is the effecting of dynamic pricing in the smart grid. This article presents an investigation into the application of M2M communications in effecting dynamic pricing in the smart grid.

Dynamic pricing of electricity is a demand-side management technique that is capable of stimulating demand response resulting in flatter load curves. Dynamic pricing policies are preferred over static pricing as these are reflective of the supply and demand imbalance which is a major problem for current power providers. The current load balancing schemes are noted not to involve the end users. However, use of pricing as a load balancing mechanism makes consumers to be more conscious of their power utilization at different times of the day and gives them a much bigger say and control in determining their own power prices. There are several dynamic pricing schemes each of which can be a suitable policy depending on the market.

M2M communications enables M2M components to be interconnected, networked, and controllable remotely, with low-cost, scalable, and reliable technologies. Because of the M2M features, a large number of metering devices can monitor and record their energy consumption and automatically report back to the utility control room through M2M communications. Automation brought about by the M2M communications helps customers to respond quickly to changes in prices while also reducing the human inefficiencies due to manual scheduling of tasks. M2M communications also enable the transmissions of usage patterns that are very vital in planning the network and also in carrying out maintenance of the grid. This contributes to achieving a balance between power generation and usage to significantly improve the power quality and efficiency of the electrical grid.

The real-time price-based coupling between supply and demand, helps in peak load reduction. However, it is noted that further research on electricity market segmentation is required to enable better implementation of dynamic prices. The benefits of M2M communications in the smart grid far outweigh the risks and this calls for wider adoption of M2M communications in the smart grid.

IX. BIBLIOGRAPHY

[1] Z. M. Fadlullah, M. M. Fouda, N. Kato, A. Takeuchi, N. Iwasaki and Y. Nozaki, "Towards Intelligent Machine-to-Machine Communications in Smart Grid," *IEEE Communications Magazine* , vol. 49, no. 4, pp. 60-65, April 2011.

[2] Y. Zhang, R. Yu, R. Yu, Y. Liu, M. Nekovee and S. Gjessing, "Cognitive Machine-to-Machine Communications: Visions and Potentials for the Smart Grid," *IEEE Network,* vol. 26, no. 3, pp. 6 - 13 , May/June 2012.

[3] Y. E. Kethsiyal, M. A. Pugazhendi, V. K. Kumar and J. Nerenjana, "Optimistic Technique for Smart Grid Infrastructure Using Sensor Network," *International Journal of Innovative Research & Development,* vol. 3, no. 2, pp. 114-119, February, 2014.

[4] S. K. Tan, M. Sooriyabandara and Z. Fan, "M2M Communications in the Smart Grid: Applications, Standards, Enabling Technologies, and Research Challenges," *International Journal of Digital Multimedia Broadcasting*, vol. 2011, pp. 1-8, 2011.

[5] Z. Fan, P. Kulkarni, S. Gormus, C. Efthymiou, G. Kalogridis, M. Sooriyabandara, Z. Zhu, S. Lambotharan and W. H. Chin, "Smart Grid Communications: Overview of Research Challenges, Solutions, and Standardization Activities," *IEEE Communications Surveys & Tutorials* , vol. 15, no. 1, pp. 21-38, 2013.

[6] J. Huang, H. Wang, Y. Qian and C. Wang, "Priority-Based Traffic Scheduling and Utility Optimization for Cognitive Radio Communication Infrastructure-Based Smart Grid," *IEEE Transactions on Smart Grid,* vol. 4, no. 1, pp. 78-86, March 2013.

[7] J. Wan, M. Chen, F. Xia, D. Li and K. Zhou, "From Machine-to-Machine Communications 1. towards Cyber-Physical Systems," *Computer Science and Information Systems*, vol. 10, no. 3, pp. 1105-1128, June 2013.

[8] R. R. Mohassel, A. Fung, F. Mohammadi and K. Raahemifar, "A survey on Advanced Metering Infrastructure," *International Journal of Electrical Power & Energy Systems,* vol. 63, pp. 473-484, December 2014.

[9] Y. Zhang, R. Yu, M. Nekovee, Y. Liu, S. Xie and S. Gjessing, "Cognitive Machine-to-Machine Communications: Visions and Potentials for the Smart Grid," *IEEE Network*, vol. 26, no. 3, pp. 6 - 13, May 2012.

[10] A. Zanella, N. Bui, A. Castellani, L. Vangelista and M. Zorzi, "Internet of Things for Smart Cities," *Internet of Things Journal, IEEE,* vol. 1, no. 1, pp. 22 - 32, 2014.

[11] J. Alonso-Zarate, J. Matamoros, D. Gregoratti and M. Dohler, "Machine-to-Machine (M2M) Communications in Smart Grid," in *Smart Grid Communications and Networking*, Cambridge University Press, 2012, pp. 147-174.

[12] N. Pandey, "White Paper on “Machine-to-Machine Communication (M2M)," 2016.

[13] H.-K. Lee, D. M. Kim, Y. Hwang, S. M. Yu and S.-L. Kim, "Feasibility of cognitive machine-to-machine communication using cellular bands," *IEEE Wireless Communications,* vol. 20, no. 2, pp. 97-103, April 2013.

[14] M. Chen, J. Wan, S. Gonzalez, X. Liao and V. C. M. Leung, "A Survey of Recent Developments in Home M2M Networks," *IEEE Communications Surveys & Tutorials,* vol. 16, no. 1, pp. 98 - 114, 2014.

[15] A. Pitì, G. Verticale, C. Rottondi, A. Capone and L. L. Schiavo, "The Role of Smart Meters in Enabling Real-Time Energy Services for Households: The Italian Case," *Energies,* vol. 10, no. 2, pp. 1-25, 2017.

[16] G. Dutta and K. Mitra, "A literature review on dynamic pricing of electricity," *Journal of the Operational Research Society*, vol. 68, no. 10, p. 1131–1145, October 2017.

[17] V. Kouhdaragh, D. Tarchi, A. V. Coralli and G. E. Corazza, "Cognitive Radio Based Smart Grid Networks," in *2013 24th Tyrrhenian International Workshop on Digital Communications - Green ICT (TIWDC)*, Genoa, 2013.

[18] S. A. Khan, M. F. Sohail, S. A. Ghauri and N. Aqdas, "Cognitive radio based Smart Grid Communication Network," *Renewable and Sustainable Energy Reviews*, vol. 72, pp. 535-548, May 2017.

[19] D. Niyato, L. Xiao and P. Wang, "Machine-to-Machine Communications for Home Energy Management System in Smart Grid," *IEEE Communications Magazine,* vol. 49, no. 4, pp. 53 - 59, 2011.

[20] V. B. Misic and J. Misic, Machine-to-Machine Communications: Architectures, Technology, Standards, and Applications, CRC Press, 2014, pp. 1-348.

[21] J. Jermyn, R. P. Jover, I. Murynets, M. Istomin and S. Stolfo, "Scalability of Machine to Machine systems and the Internet of Things on LTE mobile networks," in *2015 IEEE 16th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, Boston, 2015.

[22] Y. Zhang, R. Yu, S. Xie, W. Yao, Y. Xiao and M. Guizani, "Home M2M Networks: Architectures, Standards, and QoS Improvement," *IEEE Communications Magazine,* vol. 49, no. 4, pp. 44 - 52, April 2011.

[23] G. Wu, S. Talwar, K. Johnsson, N. Himayat and K. D. Johnson, "M2M: From Mobile to Embedded Internet,"

IEEE Communications Magazine, vol. 49, no. 4, pp. 36-43, April 2011.

[24] C. Xin and M. Song, Spectrum Sharing for Wireless Communications, Springer, 2015, pp. 1-15.

[25] G. López, P. Moura, J. I. Moreno and J. M. Camacho, "Multi-Faceted Assessment of a Wireless Communications Infrastructure for the Green Neighborhoods of the Smart Grid," *Energies*, vol. 7, pp. 3453-3483, 2014.

[26] C. Pereira and A. Aguiar, "Towards Efficient Mobile M2M Communications: Survey and Open Challenges," *Sensors*, vol. 14, no. 10, pp. 19582-19608, 2014.

Study on Wire Electrical Discharge Machining of Inconel-625

Claver Nsanzumuhire*, Bernard W. Ikua and Karanja Kabini

Abstract— Wire Electrical Discharge Machining (WEDM) is one of the non-conventional methods for machining of materials that are difficult to machine using conventional means. In WEDM, the optimum machining process parameters are governed mainly by the properties of workpiece material. Determination of these parameters presents a challenge to machining process planners. Improperly selected parameters for WEDM may result in low Material Removal Rate (MRR), short-circuiting of wire, wire breakage and poor surface finish. In this paper, the effect of Pulse-on time, Gap voltage and Wire Feed Rate on Surface Roughness (SR) and Material Removal Rate in WEDM of Inconel-625 is analyzed. Experiments are carried out for a wide range of input parameters and analysis is carried out on their effects on machining performance. The results show that both MRR and SR increase with increase in Pulse-on time. Also, the best surface finish (Ra = 3.5 μm) is obtained when the Pulse-on time is below 15 μ sec. The maximum MRR is obtained at a Pulse-on time ranging between 20-25 μsec.

Keywords—Gap Voltage, Material Removal Rate, Pulse-on time, Surface Roughness, Wire Electrical Discharge Machining, Wire Feed Rate.

I. INTRODUCTION

Inconel-625 is a nickel-alloy based material which has the properties such as high fatigue strength, resistance to chloride stress, high hardness and oxidation resistance [1]. These properties make this material suitable for applications in marine, aerospace, and power generation industries [2]. Due to its very high hardness, the material is very difficult to machine by conventional machining methods such as turning, milling or grinding. For this reason, wire electrical discharge machining is one of the alternative methods used to machine the material.

The wire EDM is an electro-thermal process that utilizes a continuously travelling wire electrode to erode material from the workpiece submerged in a dielectric fluid. Debris is flushed away from the machining zone by the dielectric fluid. Tension is applied on the wire to maintain the spark gap and to avoid tapering of the cut surface. A schematic of the wire EDM process is shown in Figure 1.

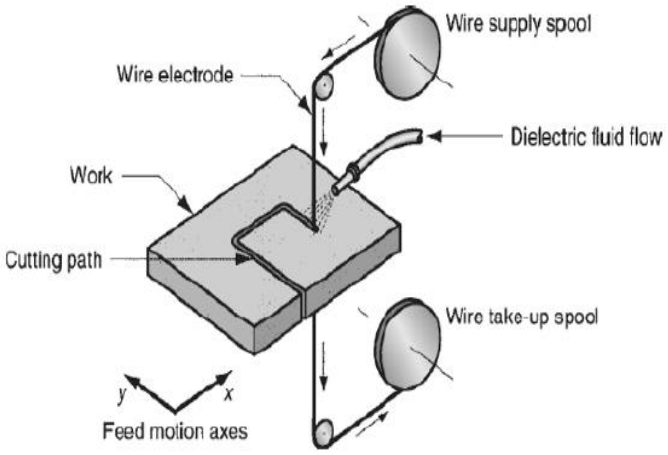


Figure 1: Schematic diagram of wire EDM [3]

II. LITTERATURE REVIEW

A great deal of research has been reported in literature, on optimization of WEDM processes. Abbas and Solomon [4] investigated the influence of two factors namely duty cycle and peak current in the wire EDM of Inconel-625. In their study, Taguchi orthogonal array was used to optimize these parameters for minimum surface roughness. It was reported that surface roughness increased with the increase in Pulse-on time and decreased with increase in Pulse-off time.

Saravan and Babu [5] conducted a study on the wire EDM of Inconel-625, in which they investigated

Claver Nsanzumuhire*, Department of Mechatronic Engineering, (+254727359364; niceracla@gmail.com)
Bernard W. Ikua, Department of Mechatronic Engineering, JKUAT (e-mail: ikua_bw@eng.jkuat.ac.ke);
Kabini Karanja, Department of Mechatronic Engineering, JKUAT (e-mail: kkabini@eng.jkuat.ac.ke)

influence of peak current, wire tension, wire diameter, Pulse-on time and Pulse-off time on material removal rate, surface finish and corner accuracy. It was reported that for optimum material removal rate, good surface finish and improved corner accuracy, the Peak current, wire tension, wire diameter, Pulse-on time and Pulse-off time should be 7 A, 1400N/mm^2, 0.25mm, 8 µs and 17µs, respectively.

Tonday and Tiga [6] investigated the effect of cutting parameters on material removal rate and surface integrity in WEDM of Inconel-625. The process parameters considered were peak current, Pulse-on time, Pulse-off time and wire tension. The study used Taguchi technique and analysis of variance (ANOVA) to analyze the influence of these parameters on output parameters. As a result, the research concluded that wire tension had no effect on both MRR and surface roughness. Also, Pulse-on time was found to have the highest impact on both the output parameters, whereas Pulse-off time had the greatest impact on MRR. It was found that the peak current had the largest impact on the surface roughness.

Singh et al., [7] conducted an experimental study on metal removal rate and surface characteristics in wire EDM of Inconel-625. In this work, the input parameters were peak current, Pulse-on time and Pulse-off time. It was reported that the increase in current and Pulse-on time increased MRR and surface roughness

Abhishek and Rahul [8] investigated the machining performance of wire EDM of Inconel-625. In their research, application of fuzzy inference systems coupled with Taguchi's philosophy was demonstrated to be an efficient method for simultaneous optimization of process parameters on MRR and SR. It was reported that the optimum values of gap voltage, peak current, Pulse-on time and flushing pressure should be 90 V, 5 A, 200 µs and 0.6 bars respectively.

From the foregoing, it can be seen that much work has been done on the effect of discharge current, Pulse-off time, duty cycle, wire tension and wire diameter, mainly using Taguchi orthogonal arrays and fuzzy inference systems. The main focus has been on how these parameters affect MRR, surface roughness and corner accuracy. However, the influence of Pulse-on time, feed rate and gap voltage were not considered. This research seeks to investigate the effect of Pulse-on time, feed rate and gap voltage on MRR and SR. Also, an algorithm for optimization of wire EDM of Inconel-625 will be developed.

III. METHODS

2.1. Equipment used in the study

Machining experiments were conducted on the AWT6S wire EDM machine. The surface roughness was measured using surface roughness tester (MITUTOYO SJ-30) and the kerf width was measured using a profile projector (Mitutoyo PJ311). The MRR was obtained from the volume of the material removed for a given machining time. Photographs of the equipment used are shown in Figures 2-4.

Figure 2: Photograph of wire EDM

Figure 3: Set up for measument of surface roughness using surface roughness tester

Figure 4: Set up for measument of gap width using profile projector

The specifications of the Wire Electrical Discharge Machine, wire EDM, surface roughness tester and profile projector are shown in table 1.

Table 1: Specifications of equipment used

Wire EDM	Model	AWT6S
	Dielectric fluid	Deionized water
	Pulse-on time range(μs)	1-24
	Pulse-off time range (μs)	1-50
	Wire feed rate range (mm/min)	1-200
	Gap voltage range (V)	1-150
Surface roughness tester	Model	PJ-301
	Stylus travel distance (mm)	4.0
	Stylus travel speed (mm/s)	0.5
Profile projector	Model	PJ-311
	Screen diameter(mm)	300
	Rotation (degrees)	360

Experiments were made using design of experiments (DOE) method and they involved machining of straight cuts. Straight cuts were chosen to allow for ease of measurement of surface roughness.

The machining experiments were conducted with workpieces of 10 mm thickness, at constant wire feed rate and gap voltage of 100.5 mm/min and 75.5 V respectively. The variable parameter was the Pulse-on time which was varied from 1-24 μsec.

The MRR was obtained as

$$MRR = V/t \quad (1)$$

Where V is the volume of material removed on machining for a duration of time t. As indicated in figure 5, it is seen that the obtained surface roughness was 3.39 μm. The upper part of the figure indicates the surface roughness parameters recorded by the machine based on the roughness profile as indicated from the bottom part of figure.

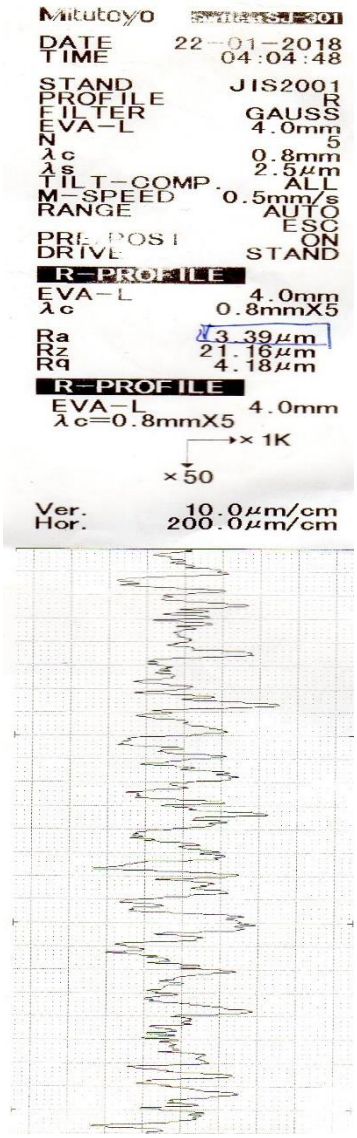


Figure 5: Typical surface roughness profile

IV. RESULTS AND DISCUSSION

Figure 5 shows a typical trace of surface roughness profile as obtained from the surface roughness tester.

The influence of Pulse-on time on SR is shown in Figure 6. It is seen in this figure that the SR is hardly affected by the Pulse-on time for up to Pulse-on time of 15 μs, above which the SR begins to increase with increase in Pulse-on time. This means that holding other

factors constant, for good surface finish, the Pulse-on time should not be above 15 µs.

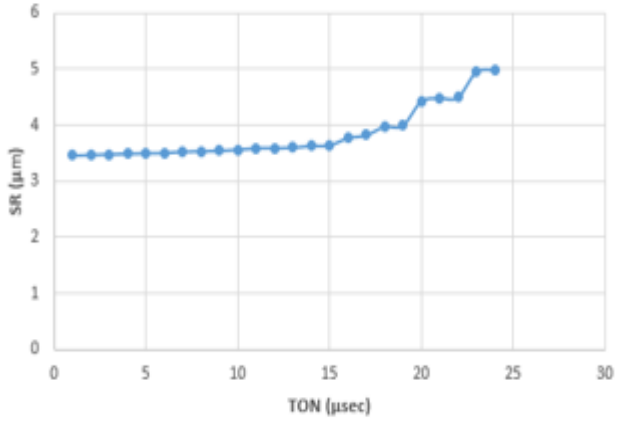


Figure 6: Influence of Pulse-on time on SR

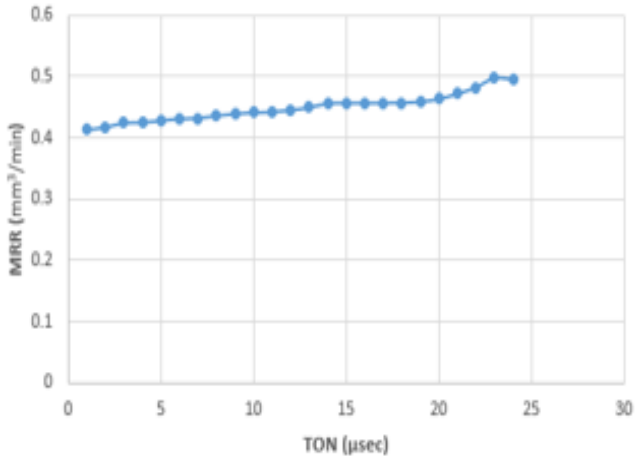


Figure 7: Influence of Pulse-on time on MRR

Figure 7 shows the influence of Pulse-on time on MRR. It is seen in this figure that the MRR increases gradually with increase in pulse-on time. This shows that to maximize on MRR, the Pulse-on time must be high. This however, would compromise on the SR, as was seen in Figure 6. Therefore, it is important to obtain an optimum value of Pulse-on time, which will give an acceptable SR and MRR.

Towards this, an algorithm for optimization of the WEDM process is necessary.

V. CONCLUSION

The effect of Pulse on-time on response variables such as material removal rate and surface roughness has been presented. It was established that both MRR and SR increase with the increase in Pulse-on time. However, the required higher MRR is compromised by the unexpected high SR. As recommendations, the effect of other machining parameters such as gap voltage and wire feed rate on the machining process should be investigated as well. An optimization algorithm should also be developed to optimize the machining process.

REFERENCES

[1] Special metals PCC Company, "EDM network," Special Metals Corporation, 2013.

[2] E. Oberg, F.D. Jones, H.L. Holton, and H.H., Machinery's Handbook. Industrial Presslnc, 1996.

[3] H. Bisaria and P.Shanbilya, "Machining of metal matrix composites by EDM and its variants": A review, DAAAM International, pp. 267-282, 2015.

[4] N. Abbas and D. Solomon, "Electrical discharge machining (EDM): Selection of dielectricin machining Inconel 718," Advanced Materials Research, 2012.

[5] M. Saravanan and A.S. Babu, "Parametric optimization of increasing corner accuracyin WEDM on inconel-625," Journal of Chemical and Pharmaceutical Sciences, 2015.

[6] H.R. Tonday and A.M. Tiga, "Analysis of effects of cutting parameters of wire electrical discharge machining on material removal rate and surface integrity," National Conference on Processing and Characterization of Materials, 2016.

[7] M. Singh, H. Lal and R. Singh, "Recent development in wire EDM," Elepean Journel of Engineering and Technology, 2013.

[8] K. Abhishek and Rahul, "Machining performance optimization for electro discharge machining of Inconel 601, 625, 718 and 825," The Brazilian Society of Mechanical Sciences and Engineering, 2016.

Dynamic Soil Moisture Control System for Irrigation Using GSM

Rabiu Aminu and Deepthi Raveendrababu Sugathakumari

***Abstract*—**Agriculture plays a significant role in economic development for the African continent. For example, 45% of the Kenyan revenue is generated from Agriculture. In Africa, agricultural activities are primarily dependents on rain which is an insufficient source and thus the need for irrigation. In order to enhance water utilization efficiency and increase crops productivity, a new system called **Dynamic Soil Moisture Control System (DSMCS)** for irrigation using Global System for Mobile Communication (GSM) technology is introduced. The proposed system uses PIC16F877 microcontroller, dc motor, soil moisture and water level sensor to monitor any changes in the soil moisture content in the farm and water level in the irrigation tank. The GSM modem is configured using Attention (AT) commands to enable a real-time monitoring of the farm remotely. All the status of the farm are dynamically monitored using DSMCS and send to the farmer through short messages (SMS), which resolves the distance and range problem. From the results obtained, the proposed system ensures adequate irrigation to plants based on soil moisture content and thus minimized water wastage.

Keywords— DC motor, GSM modem, Microcontroller, Soil moisture sensor, SMS messages water level sensor.

I. INTRODUCTION

THE significant increasing demand for food has necessitated the rapid improvement in food production technology. The major source of income for most African countries is agriculture,[1], [2]. However, majority of these countries depend on rain for agricultural activities. This high dependency on rain reduces agricultural productivity in the continent. Only few farmers in the continent practice irrigation as an alternative way of supplying water to the crops. Most of these farmers use manual and traditional irrigation methods to provide water to their crops on regular basis. These include drip irrigation, ditch irrigation, sprinkler system etc. However, the manual and traditional irrigation techniques cannot ensure a stable supply of water to the crops, they consume time and are ineffective. Due to the lack of water management in the manual irrigation techniques, sometime the field become dry and sometime flooded with excess water. This might lead to the significant waste of water which is in the red alarm in countries like Kenya, Nigeria etc. Other resources such as man power, energy and time are also wasted in the process. Thus, the automated irrigation technique needs to be designed to bring an advanced of technology into the area of agriculture which plays a significant role in the economic development of the African continent. This technique used soil moisture sensor to detect the moisture content of the soil and sends signal to the microcontroller and accordingly irrigates crops. This paper presents a dynamic soil moisture control system (DSMCS) for irrigation using Global System for Mobile Communication (GSM). The main objective is to provide adequate water to crops and at the same time minimize the water wastage by monitoring soil moisture changes. Soil moisture sensor measures changes in the moisture level of the soil and sends signal to the microcontroller and accordingly communicates to the user through GSM network. Water level sensor is added to this project to monitor the amount of water in the irrigation tank and sends the information to the user.

The rest of the paper is organized as follows. Related works are summarized in section II. The working principle of the proposed system is described in section III, which includes the system block diagram, block diagram description, and components used. In section IV the system design is discussed. Experimental results are presented in section V. Finally, conclusion and recommendation are given in section VI.

II. RELATED WORKS

Over the years, several semi-automated and automated irrigation control systems have been proposed to address water wastage and increase crops productivity. The most widely used irrigation control systems use timers, controllers and switches to irrigate crops for a specific interval of time regardless of soil moisture content. ZigBee, Bluetooth and Wi-Fi are employed in this class of irrigation control systems. These systems have advantages over the manual irrigation techniques. The major drawbacks of these systems are limited range and lack of consideration for the soil moisture content of the farmland. However, soil moisture content has crucial impact on crops growth, reproduction and photosynthesis. Thus, getting information about soil moisture content accurately, effectively and timely plays an important role in guiding water saving irrigation [3]. The breakthrough of GSM technology and soil moisture sensors increased research interest in the area of remote control systems. With the use of GSM technology, the

Rabiu Aminu, Department of Electrical Engineering, Pan African University Institute of Basic Sciences Technology and Innovation, JKUAT, +254796217162 (*email:rabiuaminu2010@gmail.com).*

Deepthi, School of Electrical and Electronics Engineering Linton University College, KTG Education Group, 12,71700 Mantin, Negeri Sembilan Malaysia *(email: deepthisankar.ssd@gmail.com)*

user and control system can communicate by sending and receiving Short Message Service (SMS) messages. Based on this technology many automated systems have been proposed, such as a remote monitoring through mobile phone using spoken commands [4], SMS controlled irrigation system with moisture sensor [5], internet based wireless home automation system for malfunctioning devices [6], a remotely controlled home automation system [7], etc.

Although, several irrigation control systems have been developed in the past years, most of them are semi-automated and do not use soil moisture sensor. Few automated irrigation systems that used moisture sensor can be found in [8], [9], [10], [11], and [12]. In general, GSM based automated irrigation control system has three important parts. The soil moisture sensor, which senses any change in the moisture level of the farm and sends the information to the control unit. The second part is the control unit. The microcontroller is the most widely used device for this purpose. It is the brain of the system, programmed to receive information from the sensing unit and communicates with the user. The microcontroller sends commands to different components of the system, handles real time data and takes timely decisions. The third part is the GSM modem, which is interfaced with the microcontroller using set of Attention (AT) commands. This device receives information from the microcontroller and sends it to the user's mobile phone through wireless communication.

To increase productivity and minimize water wastage numerous improved automated irrigation control systems have been presented recently. In [13], [14], and [15], a low cost automatic irrigation control system is introduced based on microcontroller and GSM technology. In [16], an automatic irrigation control system for smart city using Programmable Logic Controller (PLC) and Supervisory Control and Data Acquisition System (SCADA) is proposed. The system employed SCADA and fuzzy system to reduce the level of water wastage in irrigation. Namala et.al [17], presented a smart irrigation with embedded system. The system is based on the concept of Raspberry and soil moisture sensor. A solar powered automatic irrigation system on sensing moisture content using Arduino and GSM modem is introduced in [18], this system used solar energy, soil moisture sensor, GSM, and Arduino to supply adequate water to plants. In [19], Banumathi et.al presented an android based automatic irrigation system using Bayesian network with SMS voice alert. The system used Wireless Sensor Network (WSN), Arduino, and General Packet Radio Service (GPRS) module to provide water to the farms based on water level conditions. In [20], [21], an automatic irrigation systems based on Internet of Things (IoT) are proposed to minimize water wastage and human intervention in the irrigation process. Lastly, Kumar et.al [22] developed an automatic solar powered drip irrigation system using Wireless Sensor Network Technology (WSNT) by integrating Solar Photovoltaic System (SPV), Arduino microcontroller, soil moisture sensor, mobile Bluetooth, water tank, pump etc. The intension is to ensure uniform watering at right time without manual intervention and thus enhance quality and quantity of agricultural yields.

III. PROPOSED SYSTEM

Fig.1 shows the overall block diagram of Dynamic soil moisture control system for irrigation. The block diagram consists of three important parts: GSM modem, Microcontroller, and soil moisture sensor.

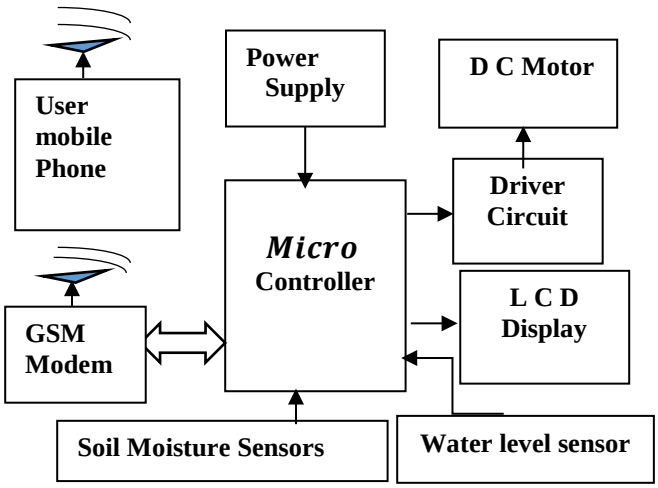


Fig.1: System block diagram

A. *Block diagram description*

The Dynamic soil moisture control system for irrigation block diagram comprises of microcontroller circuit, which acts as the central processing unit of the system. It is also interfaced with GSM modem, Liquid Crystal Display (LCD), moisture sensor and motor driver. Microcontroller is programmed with an embedded C- code to send commands to different components of the system and receives input data from soil moisture sensor and GSM modem. Through LCD a user can know the status of the field such as soil moisture content in the farmland, motor speed, and the amount of water in the tank. The GSM modem is used to describe protocols for digital cellular networks and interfaced with microcontroller to receive and send messages which indirectly control the entire system. GSM modem is controlled by a set of AT commands. Soil moisture sensor is as associated with the farm field to detect the moisture content of the soil. The data received from the sensor is sent to the microcontroller, which sends SMS messages to the user according to the received data. Water level sensor is added to this project to indicate the amount of water in the irrigation tank. The entire system is driven by 240V AC power supply.

B. *Materials*

The main materials needed for this study are described as follows:

1.1 Microcontroller

A microcontroller is a high integrated functional computer system-on-a-chip. It contains an integrated memory and programmable input/output peripherals. Microcontrollers often operate at a very low speed and consume relatively little power. Microcontroller is now available in different families such as Intel, Motorola, Peripheral Interface Controller (PIC), Atmel, and so on. The choice depends on the application and the company preferred. For this project PIC16F877 has been

selected, because of its additional features such Universal Synchronous/Asynchronous Receiver/Transmitter (USART), Analogue to Digital (A/D) converters, Pulse with Modulation (PWM) module which are very vital for this project. The PIC16F877 is a high performance modified Harvard Reduced instruction set computer (RIC) chip manufactured by microchip technology. It has 40 pins, 8k flash of program Read Only Memory (ROM), 368 byte of Random Access memory (RAM) and 256 byte of nonvolatile EEPROM memory, 33 I/O pins, 8 multiplexed A/D converters with 10 bits of resolution, PWM generator, three inbuilt timers, UART internal interrupt facilities and so on. The ROM or code ROM is used for storing program and the RAM is used for storing data. Figure 3.6 shows the pin configuration of PIC16F877 microcontroller [23].

1.2 Soil moisture sensor

A soil moisture sensor is a two terminal Integrated Circuit(IC) which uses a simple principle of conduction between to wires in a medium. This IC consists of two wires one with the 5V supply and the other one was the ground. When there is water in the field, a medium being established between the wires and conduction will take place between the wires. The voltage at the microcontroller will go low indicating the wet condition of the field. In the wet condition the motor will be in off state. If there is no water in the field, there is no conduction between the wires and the voltage at the microcontroller will go high indicating the dry condition of the field and the motor will be in ON position

1.3 GSM modem

GSM modem is used in this project to provide a two way communication between the user and the irrigation system through SMS messages. It is interfaced through the serial port of the PIC16F877 microcontroller. A set of AT commands are sent to GSM modem through its serial port, which is triggered by the microcontroller during the instances at which the interaction between the system and user are required. For this project a GSM SIM900 is used because of its special features which are very important for this study. GSM modem SIM900 is a quad-band GSM/GPRS modem engine, works on the frequencies of 850MHz, 900MHz, 1800MHz and 1900MHz and is small in size. This modem has an inbuilt RS232 level converter circuitry which allows interfacing it to the PC serial port directly. The baud rate ranges from 9600-115200 which is configurable using AT commands. It has an internal Transfer Control Protocol/Internet Protocol (TCP/IP) which makes it easy to connect with internet via GPRS. SIM900 is suitable for sending and reading SMS messages, calling as well as transfer of data application in Machine to Machine (M2M) interface. In addition, it needs only three wires that are Transmitter, Receiver and Ground (TR, RX, and GND) to interface with microcontroller and host Personal Computer (PC) [24].

1.4 DC motor

The main criteria to be considered when selecting a motor are the torque and speed of the motor. Many different motors such as servomotors, stepper motor, dc motors with and without gears are available in the market. These different motors are used according to their applications and requirements. For example, if we want high torque and precise speed we need to use servo motors, if we want to only position and if high torques not required then stepper motors are used. The motor can be selected once we know the torque and speed required for our application. A 12V DC motor is used for this project. The speed of this motor is counted in terms of rotations of the shaft per minute which is represented as RPM. The speed of this motor can easily be controlled by PWM (pulse with modulation) technique. By increasing and decreasing the duty cycle, its speed can be reduced to any desirable figure. This concept where the speed of the motor is controlled by monitoring the driven voltage ON and OFF time is called PWM technique of DC motor speed control [25]. A DC motor uses a permanent magnet to generate the magnetic field in which the armature rotates. Therefore, it can easily be modelled by using electrical circuit.

IV. SYSTEM DESIGN

The system design is divided into software and hardware design.

A. Software Design

The software used for writing the embedded C-program for implementing this project is MikroC Pro for PIC. This software makes the programming easier and the program can easily be run, built, compiled and debugged. To avoid errors when debugging the codes, the family of the needed devices is selected before writing the codes. Here PIC16F family is selected, one attribute about this software is that it support any category of PIC and for this project PIC16F877 is specifically selected.

The flowchart of the proposed system is depicted in Fig.2. The first step is to initialize the ports for LCD and motor as the outputs of the system, the PWM pin of the motor should also be initialized. The A/D converter port is declared for interfacing the moisture sensor and the serial port of the microcontroller is interfaced with GSM modem. When the system is in ON position, the soil moisture sensor will detect the soil moisture content and send the information to the microcontroller. The level of the moisture content will be displayed on LCD while at the same time sending a message to the farmer mobile phone about the status of the field (soil moisture is high or medium or low). Based on the received data, the farmer can send message to rotate the motor at the required speed. For example, the farmer can send message to rotate the motor at full speed if the field is dry, medium speed if the field has little moisture and to stop the rotation of the motor when the moisture is high. Since the process is continuous, after displaying the first outputs the microcontroller will continuously check the soil moisture content and send SMS message to the user to rotate the motor according to the received data. Anytime the soil moisture content changes, the system will communicate to the farmer instantly.

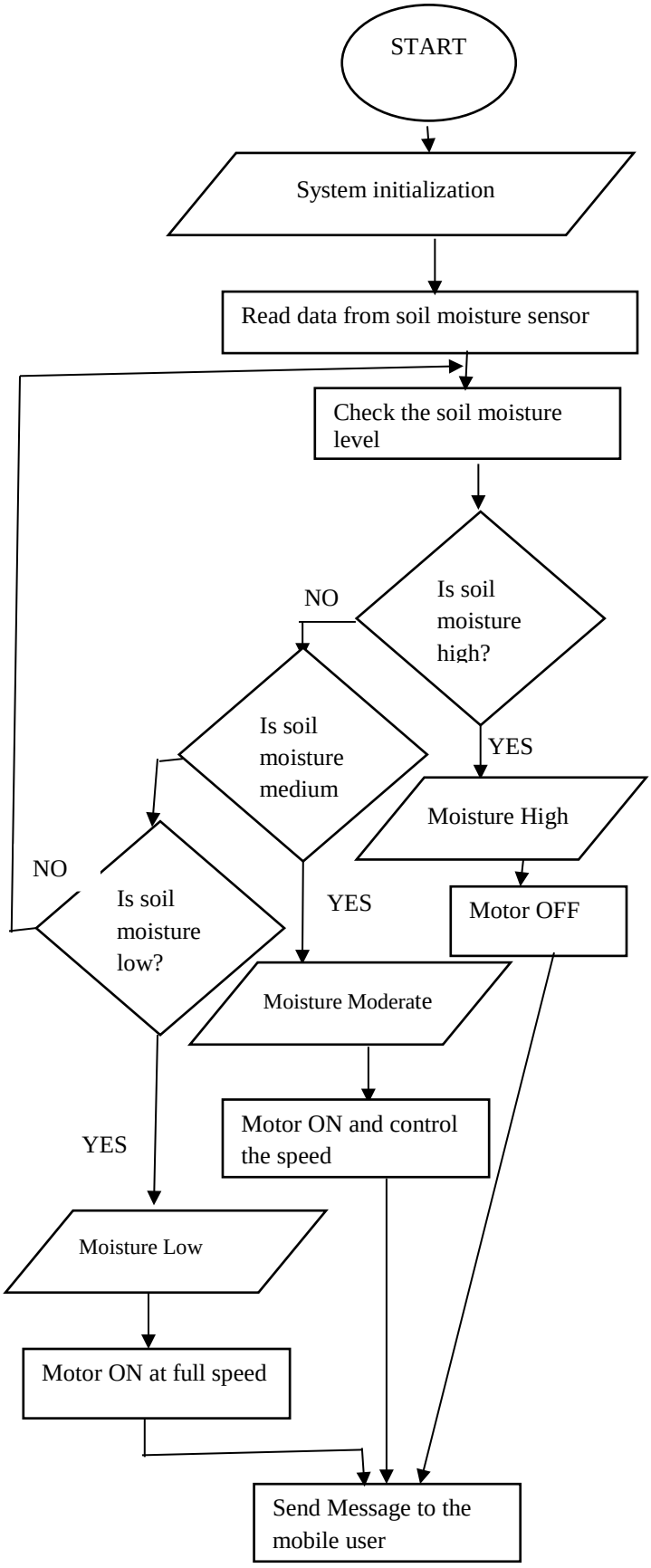


Fig.2: Flowchart of the proposed system

B. Hardware Design

The hardware implementation process involves schematic circuit design, breadboard testing, Printed Circuit Board (PCB) layout design, etching and finally system prototype testing.

1.1 Simulations diagram

The overall simulation diagram of Dynamic soil moisture control system for irrigation using GSM is shown in Fig.3.

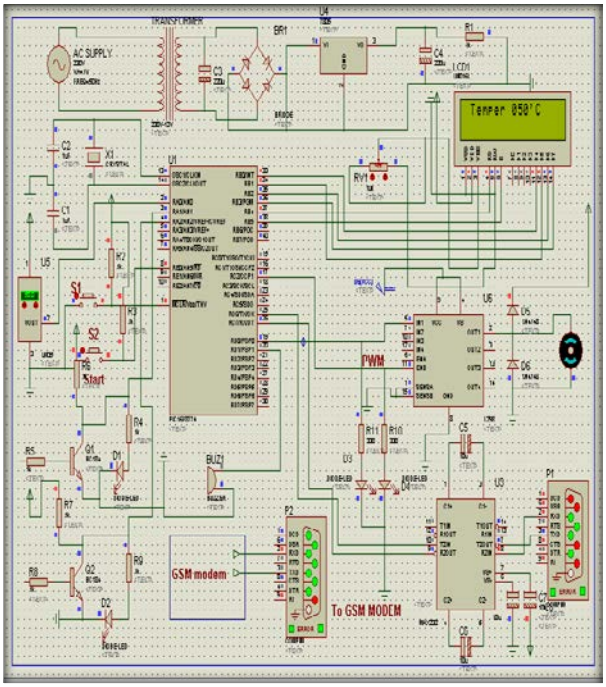


Fig.3: Overall system simulation diagram

1.2 AT commands testing

Before the breadboard testing, it is important to test the GSM modem to see if it can send and receive SMS message. In this project the test is carried out by connecting the GSM modem with computer via RS232 cable and hyper terminal software is used for verifying its functionality as shown in the Fig.4.

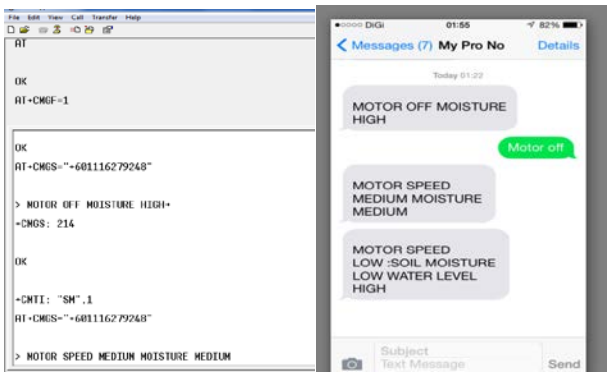


Fig.4: AT commands testing

1.2 Breadboard testing

It is important before the PCB design, to build the circuit on the breadboard to test how the circuit behaves prior to the implementation. Fig.5 shows the breadboard connection of the system.

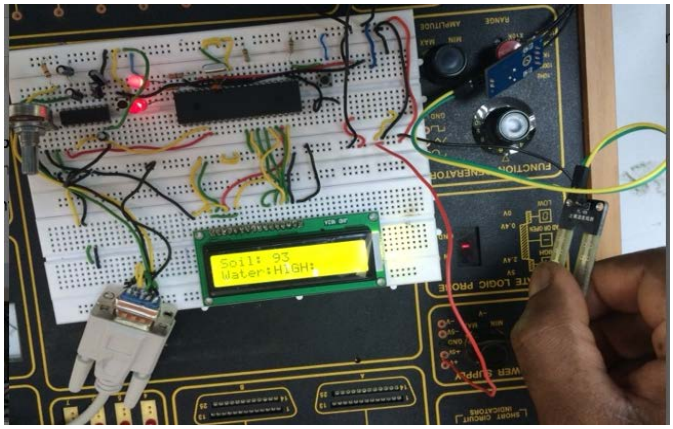


Fig.5: Breadboard connection of the system

V. EXPERIMENTAL RESULTS AND DISCUSSION

Experiment is conducted to determine the relationship between the analogue voltage, duty cycle and speed of the DC motor. The data is collected at various speed levels and corresponding voltage readings were measured using multimeter. The hardware performance is tested at various soil moisture levels. The motor speed is also recorded at high speed, medium speed and low speed.

A. Microcontroller input and output voltage

Fig.6 and Fig.7 show the voltage regulator input and output voltage measured using multimeter.

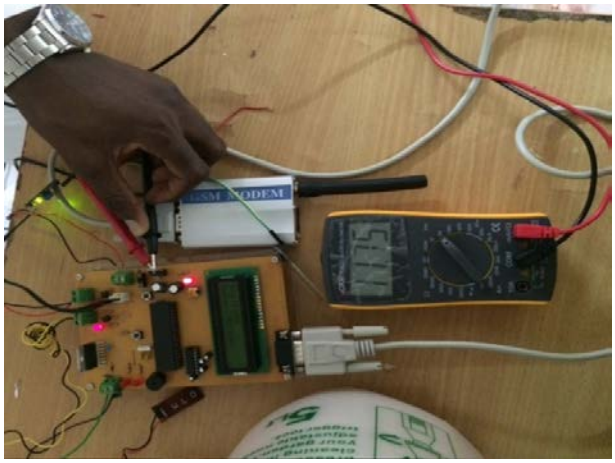


Fig.6: Voltage regulator input

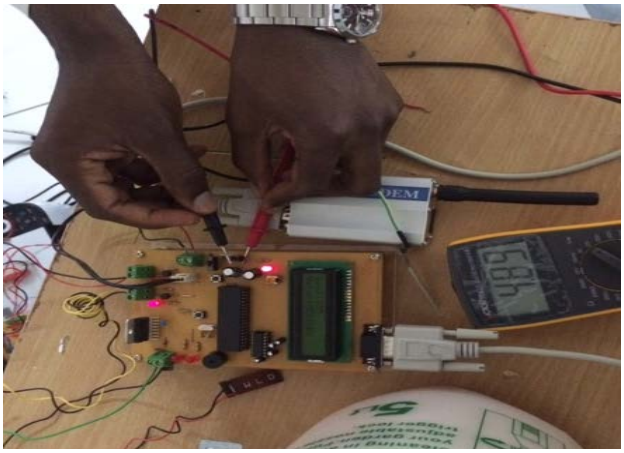


Fig.7: Voltage regulator output

Table 1: Voltage regulator LM7805 input and output

SNO	Parameter	Voltage (V)	Current (A)
1	Input	11.75	1.95
2	Output	4.89	0.3

As shown in table 1, the input current from the supply voltage is 1.95A and the corresponding output current is recorded to be 300mA. This means that the voltage regulator ensures 5V constant voltage is supplied to the microcontroller irrespective of the supply voltage fluctuation. The output current from the microcontroller is not enough to derive the dc motor. Thus, a motor driver is required to boost the current to the motor requirement.

B. DC motor driver input and output voltage

Since, the microcontroller output voltage cannot drive the dc motor due to the high current drawn by the motor when starting H-bridge driver is used. This will receive the microcontroller output as input and amplify the output current up to 2.2A which is enough to drive the motor as shown in the table 2.

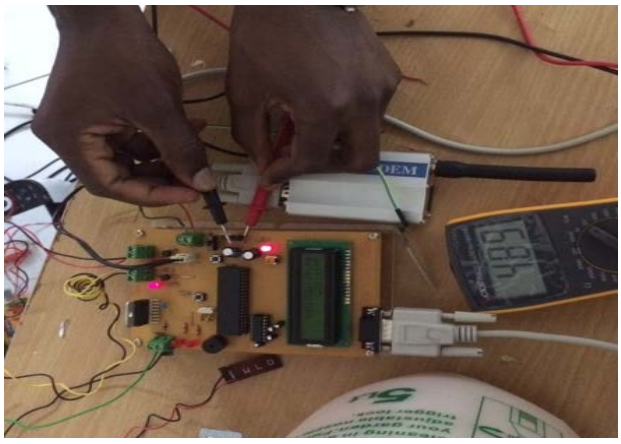


Fig.8: Motor driver input voltage

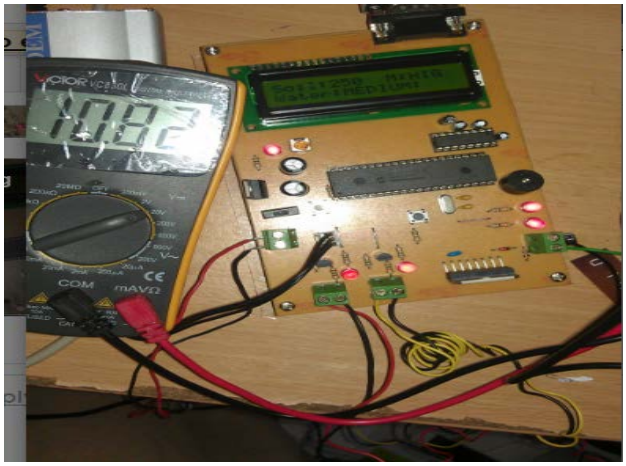


Fig.9: Motor driver output voltage

Table 2: DC motor driver input and output

S/NO	Parameters	Voltage (V)	Current (A)
1	Input	4.89	0.3
2	Output	10.82	1.53

As shown in table 2, whenever the output from the microcontroller is 0V the transistor inside the motor driver circuit will act as an open switch. The purpose is to set the driver output voltage to 10.82V. Similarly, if the output from the microcontroller is 4.89V the transistor acts as a closed switch hence 0V will appear across the driver circuit. This concept makes the motor driver to have enough current to drive the motor

C. Moisture level measurement with respect to voltage

The speed of the dc motor is measured using tachometer and the corresponding voltage at low, medium and high moisture level is recorded using voltmeter. The result is tabulated in the table 3.

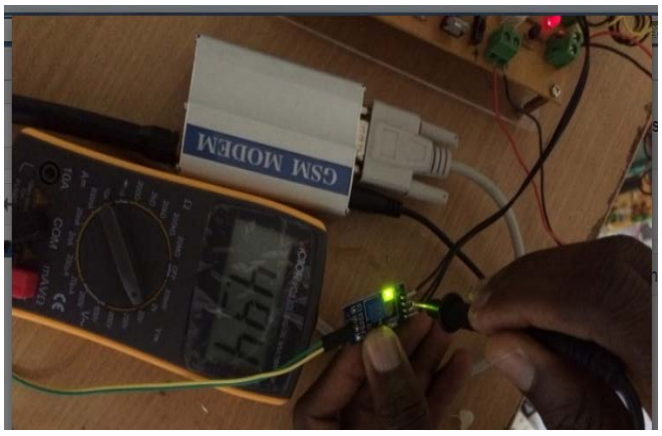


Fig.10: Voltage when soil moisture is high

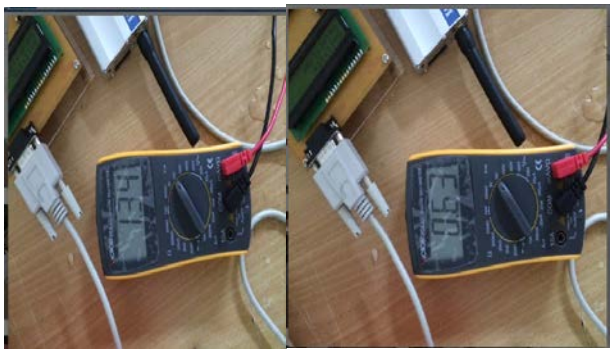


Fig.11: Voltage when soil moisture is medium and Low

Table 3: soil moisture level with voltage

S/NO	Parameters	Voltage (V)	Count Value	Status
1	High moisture	0-0.63	0-99	Moisture level high
2	Medium moisture	1.34-0.6	100-199	Moisture level Medium
3	Dried condition	4.94	200-255	Moisture level Low

A shown in Fig.10 and Fig.11, the soil moisture sensor shows a count value of 200-255 when the voltage is 4.94V, 100-199 for 1.34V and when the voltage is 0-0.63 the count value is between 0 -99 as recorded in table 3.

As shown in table 1 to table 3, the moisture level increases with decrease in voltage.

C. System prototype operation

The system prototype is depicted in Fig.13

Fig.12: System prototype

Fig.13 to Fig.15 shows the operation of the system under various conditions, Fig.13 shows the speed of the motor when the moisture is low. Fig.14 shows the motor speed when the soil

moisture is medium while Fig.15 shows the speed when the moisture is high and the motor running at high speed.

Fig.13: Moisture low motor running at High speed

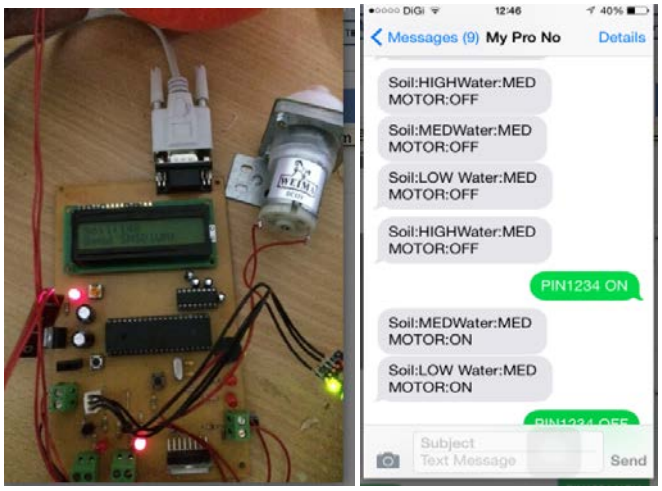


Fig.14: Moisture medium motor running at medium speed

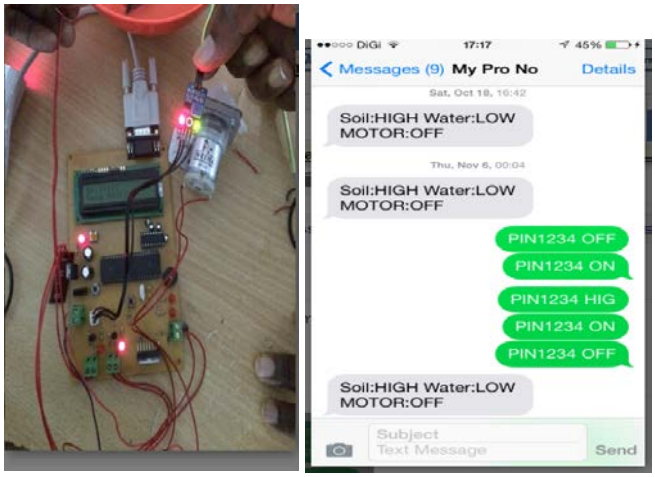


Fig.15: Moisture high motor running at low speed

VI. Conclusion

A Dynamic soil moisture control system for irrigation using GSM was successfully designed and implemented in this paper. The system notifies the user any changes in soil moisture content in the farmland and water level in the irrigation tank through SMS messages. The proposed DSMCS also provides feature to the farmer to select the speed at which the system should operate such as high, medium or low speed depends on the condition of the farm. Thus, the DSMCS proved that, the irrigation system can be controlled remotely and operated from far places. In addition, the developed system ensures effective irrigation to plants and thus minimized water wastage which is on red alarm in the African continent.

For future enhancement of this work, the system could be designed to use solar energy as its power supply which is one of the non-conventional energy sources, especially for some areas where the supply of electricity is not constant. Furthermore, instead of supplying water manually to the irrigation tank, the system can be modified in such a way that the irrigation water can be driven from the rain water harvesting (RWH) tanks installed on the farms. Moreover, the system can be enhanced so that the user can send an SMS message to the system to refill the irrigation tank when it is low. By rotating the motor to suck the water from the ground, this will save time and energy. The Dynamic soil moisture control system for irrigation using GSM could further be improved, so that if the water level reaches to the danger level, the motor will automatically start to ensure proper water in the farm field. The system could also be improved to use voice message so that even disabled community people especially those who lost their eyesight can operate the system.

Acknowledgement

The authors would like to thank Petroleum Technology Development Fund (PTDF) for sporting this work. Our appreciation to all the lecturers of the School of Electrical and Electronics Engineering Linton University College, KTG Education Group.

References

[1] X. Diao, P. Hazell, and J. Thurlow, “The Role of Agriculture in African Development,” *World Development*, vol. xx, no. x, 2010.

[2] R. E. Clute and R. E. Clute, “The Role of Agriculture in African Development",*African Study Association* , vol. 25, no. 4, pp. 1–20, Jun.2014.

[3] C. Yu, Y. Wu, and Y. Yu, “Intelligent irrigation system based on fuzzy control,” *2010 International Conference on Artificial Inteligence and Conputational Inteligence (AICI)* , vol. 3, pp. 519–521, 2010.

[4] N. P. Jawarkar and V. Ahmed, “Micro-controller based Remote Monitoring using Mobile through Spoken Commands,” *Journal of Networks*, vol. 3, no. 2, pp. 58–63, Feb. 2008.

[5] A. H. Atodaria V. H. Atodaria, V, M. Tailor A. M. Tailor and Z. N. Shah Z. N. Shah, “SMS Controlled Irrigation System with Moisture Sensors,” *Indian*

Journal of Applied Research, vol. 3, no. ISSN-2249-555X, pp. 1–2, 2011.

[6] A. Z. Alkar and U. Buhur, "An internet based wireless home automation system for multifunctional devices," *IEEE Transaction on Consumer Electronics*, vol. 51, no. 4, pp. 1169–1173, Jul.2005.

[7] N. Singh, S. S. Bharti, R. Singh, and D. K. Singh, "Remotely controlled home automation system,"*IEEE International Conference on Advances on Engineering and Technology Research (ICAETR-2014)*, Aug. 2014.

[8] J. Uddin, S. M. T. Reza, Q. Newaz, J. Uddin, T. Islam, and J. Kim, "Automated Irrigation System Using Solar Power,"*2012 7th International Conference on Electrical and Computer Engineering Bangladesh*, pp. 228–231, Dec. 2012.

[9] A. M. Hassan, "Web-based irrigation management for open canals using Wireless Sensor Networks," *2013 IEEE Conference on Wireless Sensors Sarawak*, pp. 102–107, Dece. 2013.

[10] D. Kohanbash, A. Valada, and G. A. Kantor, "Irrigation Control Methods for Wireless Sensor Network," *American Society of Agricultural and Biological Engineering Annual Meeting*, vol. 7004, no. 12, Aug. 2012.

[11] J. Gutiérrez, J. F. Villa-medina, A. Nieto-garibay, and M. Á. Porta-gándara, "Automated Irrigation System Using a Wireless Sensor Network and GPRS Module,"*IEEE Transactions on Instrumentation and Measurement*, pp. 1–11, May, 2013.

[12] A. Anil, A. R. Thampi, P. J. M, and K. J. Shanthi, "Project HARITHA - An Automated Irrigation System for Home Gardens," *IEEE Transaction on Computer Engineering*, pp. 635–639, 2012.

[13] I. Journal, F. Technological, D. S. Dasare, P. S. Kale, P. R. Kale, and H. Mande, "Automatic irrigation control system," *International Journal for Technological Research*, vol. 4, no. 8, pp. 1250–1252, Apr. 2017.

[14] B. D. Kumar, P. Srivastava, R. Agrawal, and V. Tiwari, "Microcontroller Based Automatic Plant Irrigation System," *International Research Journal of Engineering and Technology*, vol.4, no.5, pp. 1436–1439, May, 2017.

[15] J. Ashok, M. Balaji, S. Dhinakaran, and M. D. Kumar, "A Study on Microcontroller Based Automatic Real Time Water Irrigation Management System," *Journal of Chemical and Phermaceutical Sciences*, no. 4, pp. 2013–2015, March, 2017.

[16] G. Bharathi and C. G. Prasunamba, "Automatic Irrigation System for Smart City Using PLC AND SCADA," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 2, no. 4, pp. 309–314, Aug.,2017.

[17] K. K. Namala, K. K. P. A. V, A. Math, A. Kumari, and S. Kulkarni, "Smart Irrigation with Embedded System," *2016 IEEE Bombay Section Symposium (IBSS)*, vo.1, no.2, May, 2016.

[18] S. Suman, S. Kumar, R. Sarkar, and G. Ghosh, "Solar Powered Automatic Irrigation System on Sensing Moisture Content Using Arduino and GSM," *International Journal of Advanced Research in Electronics and Communication Engineering*, vol. 6, no. 6, Jun., 2017.

[19] P. Banumathi, D. Saravanan, M. Sathiyapriya, and V. Saranya, "An Android Based Automatic Irrigation System Using Bayesian Network with SMS and Voice Alert," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*,vol. 2, no. 2, pp. 573–578, March.,2017.

[20] K. Kansara, V. Zaveri, S. Shah, S. Delwadkar, and K. Jani, "Sensor based Automated Irrigation System with IOT : A Technical Review," *nternational Journal of Computer Science and Information Technologies*, vol. 6, no. 6, pp. 5331–5333, 2015.

[21] P. Naik *et al.*, "Arduino Based Automatic Irrigation System Using IoT," *nternational Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 2, no. 3, pp. 881–886, Jun., 2017.

[22] S. Kumar, C. Sethuraman, and K. Srinivas, "Solar Powered Automatic Drip Irrigation System (SPADIS) using Wireless Sensor Network Technology," *International Research Journal of Engineering and Technology*, pp. 722–731, Jul., 2017.

[23] D. Jasio,T.Wiilmshurst, M.P.Bates, C.Hellebuyk, D. Ibrahim et.al, PIC Microcontrollers:know it all,"1st Edition",United Kingdom: Lincre House, Jordan Hill, Oxford OX28DP,2008.

[24] R. Pujar, "Difference between SIM900 and SIM900A GSM modems," *Valetron system pvt ltd*, 2015. [Online]. Available: http://www.raviyp.com/embedded/174.

[25] B. A. Thomas, S. Michahial, P. Shreeraksha, V. B. Nagvi, and M. Suresh, "Industry Based Automatic Robotic Arm," *International Journal of Engineering and Innovative Technology*, vol. 2, no. 11, pp. 42–48, May, 2013.

Assessment of Voltage Stability Indices for Determination of Power System Weak Buses.

Weldon K. Koskei [1], David K. Murage[2], Samuel Kangethe[3] and Michael J. Saulo[4]

Abstract—The ever growing population and the high level of industrialization globally has led to increase demand for electric power. However, the expansion of electric power generation as well as transmission and distribution power system has been limited by financial constraints. This has led to operation of power system close to its operating stability limits. This overloaded operation condition of power system leads to instability challenges like voltage instability due to inability of the power system to supply required reactive power. In the case that voltage stability status of a stressful system not determined and the challenges arising not addressed, this can lead to voltage collapse and total blackout of the system. Hence there is need to constantly check the status of voltage stability and determine voltage collapse proximity and plan counter corrective measures accordingly. This paper aims to compare Fast Voltage Stability Index (FVSI) and Line stability Index (L_{mn}). These line voltage stability indices indicates the proximity of power system lines to voltage instability. This will help power system planners and operators in their daily power system operations. The standard IEEE 30 bus power system was used to carry out this research and test performance of these voltage stability indices.

Keywords— Voltage Stability, Voltage Collapse, Fast Voltage Stability Index, Line Stability Index.

I. INTRODUCTION

The ever growing demand for electric power due to rising business operations as well as rapid growth of industries has led to operation of power systems closely to their stability limits. These power systems are heavily loaded due to high level of electricity consuming demand which has led to instability challenges [1], [2]. Voltage instability due to insufficient reactive power in the power system is one of the major challenges to power system operators and planners.

Voltage stability is referred to as ability of the power system to maintain steady voltages at all buses in the system under a normal operating conditions and after being subjected to a disturbance [3], [4]. Voltage instability is mainly caused by increase in the load demand and inability of the system to supply required reactive power which leads to uncontrollable voltage drops. Voltage stability is classified into large disturbance and small disturbance voltage stability. Large disturbance voltage stability is concern with maintaining voltages after a large disturbances such as faults and loss of generation in the system while Small disturbance voltage stability deals with maintaining voltages after a small disturbances such as gradual changes of the loads in the system. Voltage stability analysis have gained more interest due to severe voltage instability incidents which occurred in some countries like Japan, Belgium, USA and many parts of the world some decades ago [5], [6].

The information about how close the power system operating point is from voltage collapse is very important to system planners and operators in planning counter corrective measures accordingly. Voltage stability indices indicates the proximity of power system to voltage instability where values close to zero are considered stable but values close to 1 are considered unstable.

Voltage stability analysis using different indices was carried out in this paper and the results obtained from simulating standard IEEE 30 bus were discussed.

II. VOLTAGE STABILITY INDICES

The status of voltage stability in a power system can be determined by using different tools to estimate the closeness of a various operating point to voltage instability condition. Loading margin is one of the tools used and it involve increment of load demand until voltage collapse is reached. It uses PV and QV curves which need to be generated at every bus requiring many calculations and more time. Other voltage stability indices which are used and compared in this paper are discussed below.

A. FAST VOLTAGE STABILITY INDEX

This voltage stability index was developed by Ismail Musirin [7] by considering current equation through a transmission line in two bus system as shown in Fig 1.

Weldon K. Koskei, MSc Student, Department of Electrical and Electronic Engineering, JKUAT, (phone: +254729918244; e-mail: *wkoskei@jkuat.ac.ke*),
David K. Murage, Department of Electrical and Electronic Engineering, JKUAT, (e-mail: *dkmurage25@yahoo.com*),
Samuel Kangethe, Department of Electrical and Electronic Engineering, JKUAT, (e-mail: samuel.kangethe@jkuat.ac.ke),
Michael J. Saulo, Department of Electrical and Electronic Engineering, Technical University of Mombasa, TUM, (e-mail: michaelsaulo@tum.ac.ke),

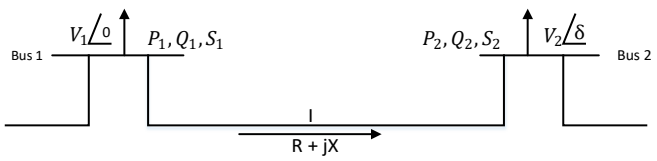


Fig 1: Two - Bus transmission line representation

Where

V_1 , V_2 is the voltages at sending and receiving buses

P_1 , Q_1 is the active and reactive powers at the sending bus

P_2 , Q_2 is the active and reactive powers at the receiving bus

S_1 , S_2 is the apparent powers at sending and receiving bus

δ is the angle difference between sending and receiving bus voltages ($\delta_2 - \delta_1$)

Taking sending bus as reference bus i.e $\delta_1 = 0$ and $\delta_2 = \delta$, the general current equation can be written as

$$I = \frac{V_1\angle 0 - V_2\angle\delta}{R+jX} \quad (1)$$

Where R is the line resistance and X is the line reactance

$$V_2^2 - \left(\frac{R}{X}\sin\delta + \cos\delta\right)V_1V_2 + \left(X + \frac{R^2}{X}\right)Q_2 = 0 \quad (2)$$

The roots for V_2 is calculated as

$$V_2 = \frac{\left(\frac{R}{X}\sin\delta+\cos\delta\right)V_1 \pm \sqrt{[(\frac{R}{X}\sin\delta+\cos\delta)V_1)]^2 - 4\left(X+\frac{R^2}{X}\right)Q_2}}{2} \quad (3)$$

The real roots of V_2 can be obtained by setting discriminant as greater than or equal to zero as

$$\frac{4Z^2Q_2X}{(V_1)^2(R\sin\delta+X\cos\delta)^2} \leq 1 \quad (4)$$

The angle difference is normally very small ($\delta \approx 0$), hence, $R\sin\delta \approx 0$ and $X\cos\delta \approx X$

By taking the symbol 'i' as the sending end and 'j' as receiving end, the fast voltage stability index can be expressed as

$$FVSI_{ij} = \frac{4Z_{ij}^2 Q_j}{V_i^2 X_{ij}} \quad (5)$$

Where

Z_{ij} - line impedance

Q_j - receiving end reactive power

V_i - sending end voltage

X_{ij} - line reactance

If the line index value is close to 0, the line is voltage stable and if it is close to 1, it is close to voltage collapse.

B. LINE STABILITY INDEX

This voltage stability index was developed by Moghavvemi [8] by considering power flow equations through a transmission line in two bus system as shown in Fig 2.

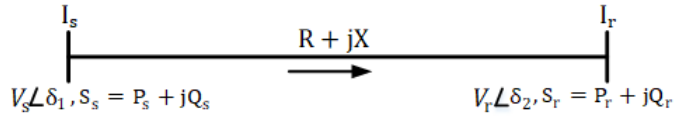


Fig 2: One-Line diagram of transmission line

Where

V_s , V_r is the voltages at sending and receiving buses

P_s , Q_s is the active and reactive powers at the sending bus

P_r , Q_r is the active and reactive powers at the receiving bus

S_s , S_r is the apparent powers at sending and receiving bus

δ is the angle difference between sending and receiving bus voltages ($\delta_2 - \delta_1$)

By applying power flow concept in the line, the power flow at the sending and receiving end can be formulated as θ - line impedance angle

$$S_s = \frac{|V_s|^2}{Z}\angle\theta - \frac{|V_s||V_r|}{Z}\angle(\theta + \delta_1 - \delta_2) \quad (6)$$

$$S_r = \frac{|V_s||V_r|}{Z}\angle(\theta - \delta_1 + \delta_2) - \frac{|V_r|^2}{Z}\angle 0 \quad (7)$$

Separating active and reactive powers from above equations will give

$$P_r = \frac{V_sV_r}{Z}\cos(\theta - \delta_1 + \delta_2) - \frac{V_r^2}{Z}\cos\theta \quad (8)$$

$$Q_r = \frac{V_sV_r}{Z}\sin(\theta - \delta_1 + \delta_2) - \frac{V_r^2}{Z}\sin\theta \quad (9)$$

Replacing $\delta_1 - \delta_2 = \delta$ in the above reactive power equation and solving for V_r

$$V_r = \frac{V_s\sin(\theta-\delta) \pm \sqrt{[(V_s\sin(\theta-\delta))]^2 - 4ZQ_r\sin\theta}}{2\sin\theta} \quad (10)$$

Substituting Z sin θ = X in the above equation

$$V_r = \frac{V_s\sin(\theta-\delta) \pm \sqrt{[(V_s\sin(\theta-\delta))]^2 - 4XQ_r}}{2\sin\theta} \quad (11)$$

To obtain real roots of V_r , the following conditions must be satisfied

$$\{[(V_s\sin(\theta-\delta))]^2 - 4XQ_r\} \geq 0 \quad (12)$$

$$\frac{4XQ_r}{[(V_s\sin(\theta-\delta))]^2} \leq 1 \quad (13)$$

By taking the symbol 'i' as the sending end and 'j' as receiving end, the Line stability index can be expressed as

$$L_{mn} = \frac{4X_{ij}Q_j}{[V_i\sin(\theta-\delta)]]^2} \quad (14)$$

Where

X_{ij} - line reactance

Q_j - receiving end reactive power

V_i - sending end voltage

θ - line impedance angle

δ - angle between sending end and receiving end voltage

If the line index value is close to 0, the line is voltage stable and if it is close to 1, it is close to voltage collapse.

III. TEST SYSTEM AND SIMULATIONS RESULTS

The voltage stability analysis was performed on standard IEEE 30 bus system. This standard test system is as shown in Fig 3. This system is composed of six generator buses, twenty four load buses and forty one interconnected transmission lines. Static loads were considered in this simulations and the system data is as shown in the appendix.

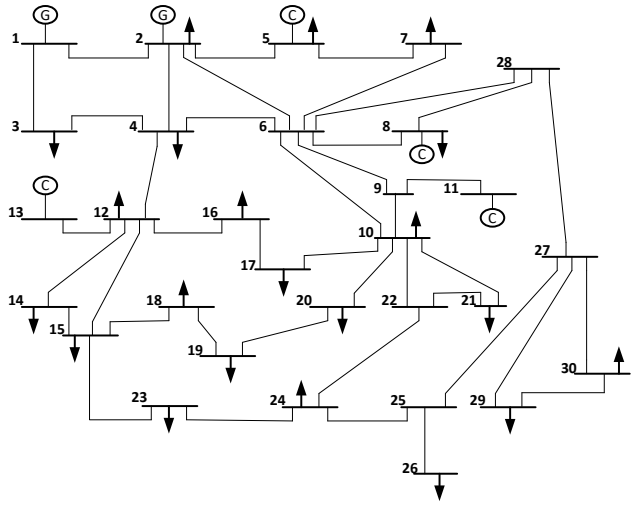


Fig 3: Standard IEEE 30 Bus Test System

The load flow simulations were carried out using MATLAB software and the results obtained were used to calculate the indices as per the formulas above. Different scenarios were considered to cater for light loading, base loading and overloading conditions and some lines were randomly chosen for analysis. Table I shows calculated indices for base loading. Table II shows the indices obtained for light loading (90% of base loading) and Table III shows indices calculated from results obtained for overloading condition (110% of base loading).

Table I: Base Loading Results

From Bus	To Bus	FVSI	L_{mn}
3	4	0.0347	0.0321
6	9	0.0125	0.0087
6	10	0.0359	0.0367
6	28	0.0138	0.0129
8	28	0.0391	0.0288
16	17	0.0156	0.0247
15	18	0.0207	0.0198
25	26	0.2537	0.2522
25	27	0.0311	0.0293
27	28	0.1207	0.1141
27	29	0.2528	0.2572
27	30	0.2622	0.2575

The values of the calculated indices in table I above generally ranges between 0.0125 and 0.2622 for both FVSI and L_{mn}. These values are closer to zero indicating that the system is stable at base loading condition. Generally the values of FVSI are slightly higher than those of L_{mn} for this scenario.

Table II: Light Loading Results

From Bus	To Bus	FVSI	L_{mn}
3	4	0.0287	0.0211
6	9	0.0118	0.0068
6	10	0.0318	0.0278
6	28	0.0104	0.0119
8	28	0.0245	0.0230
16	17	0.0142	0.0092
15	18	0.0187	0.0191
25	26	0.1491	0.1475
25	27	0.0226	0.0237
27	28	0.1182	0.1024
27	29	0.1512	0.1501
27	30	0.1594	0.1521

The results in table II were obtained by reducing both active and reactive power demand of the nominal loads at each bus. These results under light loading conditions further shows the system is stable under this scenario as the values of both FVSI and L_{mn} indices are closer to zero. This shows that the reactive power generated by the system meet the load demand. Just like base loading condition, this scenario shows that the values of FVSI are slightly higher than those of L_{mn}.

Table III: Overloading Condition Results

From Bus	To Bus	FVSI	L_{mn}
3	4	0.0621	0.0539
6	9	0.0493	0.0281
6	10	0.1028	0.0948
6	28	0.0812	0.0582
8	28	0.0901	0.0938
16	17	0.0625	0.0611
15	18	0.0862	0.0749
25	26	**0.7122**	**0.7037**
25	27	0.2404	0.2476
27	28	0.2870	0.3146
27	29	**0.8331**	**0.8299**
27	30	**0.8791**	**0.8701**

The results in table III were obtained by increasing both active and reactive power demand of the base loads at each bus. The results from some branches show that the system is stressed under this scenario. The indices values for line 25-26, 27-29 and 27-30 are shown to be closer to one indicating that they are heavily loaded and can cause a voltage collapse of the system. The generators under this operation condition are not capable to supply reactive power required by the loads. Hence the buses connected to these branches are identified as weakest points in the system. The values of both FVSI and L_{mn} show that bus 26,

29 and 30 are weak buses in this system. The highest index is obtained at bus 30 indicating that it is the weakest bus in the whole system.

The results obtained from all simulation scenarios generally show that the values of FVSI are higher as compared to those of L_{mn}. This shows that FVSI index is most sensitive in identifying weakest points in the power system as compared to L_{mn} index.

IV. Conclusion

Voltage stability index can be used to identify the weak points in the power system. In this paper, two voltage stability indices were compared under different loading conditions of the system. From the results obtained and analyses, FVSI index was found to be more sensitive in identifying weak areas of a power system as compared to L_{mn} index. Identification of these weak points can help in placement of reactive power compensation systems in the power system to avoid voltage collapse.

References

[1] Wu, Y.-K., A novel algorithm for ATC calculations and applications in deregulated electricity markets. International Journal of Electrical Power & Energy Systems, 2007.29(10):p.810-821.

[2] Savulescu, S.C., Real-time stability assessment in modern power system control centers. Vol. 42.2009:John Wiley & Sons.

[3] P. Kundur, Power System Stability and Control, McGraw-Hill, 1993

[4] C. W. Taylor, Power System Voltage Stability McGraw-Hill, 1994.

[5] "Voltage Stability of Power Systems: Concepts, Analytical Tools and Industry Experience", IEEE Committee Vol.I EEE/PES 93TH0358-2-PWR 1990.

[6] K. Takahashi and Y. Nomura "The Power System Failure on July 23rd 1987 in Tokyo" CIGRE SC-37 Meeting 37.87(JP)07(E) 1987.

[7] T. K. A. R. Ismail Musirin, "Novel Fast Voltage Stability Index (FVSI) for Voltage Stability Analysis in Power Transmission System," in *Student Conference on Research and Development Proceeding,*, 2002, pp. 265–268.

[8] M. Moghavvemi, "New method for indicating voltage stability in power system" Proceedings of IEEE International Conference on Power Engineering, Singapore, lPEC, 1997, pp. 223-227.

Appendix

IEEE 30 BUS SYSTEM DATA

Table A1 : Bus Data

Bus No	type	Pd	Qd	Vm	Va	Vmax	Vmin
1	3	0	0	1.06	0	1.06	0.94
2	2	21.7	12.7	1.043	-5.48	1.06	0.94
3	1	2.4	1.2	1.021	-7.96	1.06	0.94
4	1	7.6	1.6	1.012	-9.62	1.06	0.94
5	2	94.2	19	1.01	-14.37	1.06	0.94
6	1	0	0	1.01	-11.34	1.06	0.94
7	1	22.8	10.9	1.002	-13.12	1.06	0.94
8	2	30	30	1.01	-12.1	1.06	0.94
9	1	0	0	1.051	-14.38	1.06	0.94
10	1	5.8	2	1.045	-15.97	1.06	0.94
11	2	0	0	1.082	-14.39	1.06	0.94
12	1	11.2	7.5	1.057	-15.24	1.06	0.94
13	2	0	0	1.071	-15.24	1.06	0.94
14	1	6.2	1.6	1.042	-16.13	1.06	0.94
15	1	8.2	2.5	1.038	-16.22	1.06	0.94
16	1	3.5	1.8	1.045	-15.83	1.06	0.94
17	1	9	5.8	1.04	-16.14	1.06	0.94
18	1	3.2	0.9	1.028	-16.82	1.06	0.94
19	1	9.5	3.4	1.026	-17	1.06	0.94
20	1	2.2	0.7	1.03	-16.8	1.06	0.94
21	1	17.5	11.2	1.033	-16.42	1.06	0.94
22	1	0	0	1.033	-16.41	1.06	0.94
23	1	3.2	1.6	1.027	-16.61	1.06	0.94
24	1	8.7	6.7	1.021	-16.78	1.06	0.94
25	1	0	0	1.017	-16.35	1.06	0.94
26	1	3.5	2.3	1	-16.77	1.06	0.94
27	1	0	0	1.023	-15.82	1.06	0.94
28	1	0	0	1.007	-11.97	1.06	0.94
29	1	2.4	0.9	1.003	-17.06	1.06	0.94
30	1	10.6	1.9	0.992	-17.94	1.06	0.94

Table A2: Generator Data

Bus No	Pg	Qg	Qmax	Qmin	Vg	mBase	Pmax	Pmin
1	260.2	-16.1	10	0	1.06	100	360.2	0
2	40	50	50	-40	1.045	100	140	0
5	0	37	40	-40	1.01	100	100	0
8	0	37.3	40	-10	1.01	100	100	0
11	0	16.2	24	-6	1.082	100	100	0
13	0	10.6	24	-6	1.071	100	100	0

Table A3: Line Data

Line No	From bus	To bus	r	x	b	ratio
1	1	2	0.0192	0.0575	0.0528	0
2	1	3	0.0452	0.1652	0.0408	0
3	2	4	0.057	0.1737	0.0368	0
4	3	4	0.0132	0.0379	0.0084	0
5	2	5	0.0472	0.1983	0.0418	0
6	2	6	0.0581	0.1763	0.0374	0
7	4	6	0.0119	0.0414	0.009	0
8	5	7	0.046	0.116	0.0204	0
9	6	7	0.0267	0.082	0.017	0
10	6	8	0.012	0.042	0.009	0
11	6	9	0	0.208	0	0.978
12	6	10	0	0.556	0	0.969
13	9	11	0	0.208	0	0
14	9	10	0	0.11	0	0
15	4	12	0	0.256	0	0.932
16	12	13	0	0.14	0	0
17	12	14	0.1231	0.2559	0	0
18	12	15	0.0662	0.1304	0	0
19	12	16	0.0945	0.1987	0	0
20	14	15	0.221	0.1997	0	0
21	16	17	0.0524	0.1923	0	0
22	15	18	0.1073	0.2185	0	0
23	18	19	0.0639	0.1292	0	0
24	19	20	0.034	0.068	0	0
25	10	20	0.0936	0.209	0	0
26	10	17	0.0324	0.0845	0	0
27	10	21	0.0348	0.0749	0	0
28	10	22	0.0727	0.1499	0	0
29	21	22	0.0116	0.0236	0	0
30	15	23	0.1	0.202	0	0
31	22	24	0.115	0.179	0	0
32	23	24	0.132	0.27	0	0
33	24	25	0.1885	0.3292	0	0
34	25	26	0.2544	0.38	0	0
35	25	27	0.1093	0.2087	0	0
36	28	27	0	0.396	0	0.968
37	27	29	0.2198	0.4153	0	0
38	27	30	0.3202	0.6027	0	0
39	29	30	0.2399	0.4533	0	0
40	8	28	0.0636	0.2	0.0428	0
41	6	28	0.0169	0.0599	0.013	0

Effects of different fine aggregate on concrete strength

Kiambigi, MAINA [1, 3]; GWAYA, A.O [1] and KOTENG, D.O [2]

Abstract— Fine aggregate has been extensively used in the construction industry as a key component of concrete production. Although river sand is one of the major sources of fine aggregate, different sources exhibit different properties by virtue of the geological formation of the drainage basin. Further, the use of river sand as the source of fine aggregate has resulted in over-exploitation leading to depletion and environmental degradation. This has led to exploration of alternative sources to safeguard depletion and reduce the negative impacts on the environment. This research was conducted on a variety of river sands and alternative fine aggregates to assess their suitability for concrete manufacture. A quantitative experimental approach was adopted to test the Physical, chemical and mineralogical properties of fine aggregates sourced from Machakos, Mwingi, Naivasha and Kajiado and the resultant concrete strength after 7, 14 and 28 days recorded. The fineness modulus of all the material samples ranged from 1.92 to 3.66, specific gravity 1.73 to 2.27 and silt content 2.06% to 11.9%. All the samples fell within the overall grading envelope. The silicon dioxide concentration ranged from 65% to 80%, Aluminium oxide 9% to 19% and Calcium oxide 1.3% to 2.5%. Machakos sand had the highest Silicon dioxide and calcium oxide concentration of 80% and 2.5% respectively, while quarry dust had the highest aluminium oxide concentration of 19%. It was observed that concrete produced from natural river sand obtained from Mwingi , Kajiado and Machakos achieved strengths of 41.899N/mm^2, 37.173N/mm^2 and 33.645N/mm^2 respectively comparative to 30 N/mm^2 target characteristic strength after 28 days. On the other hand, concrete produced using fine aggregates obtained from Mlolongo rock sand, Naivasha sand and Mlolongo Quarry dust achieved strengths of 28.682 N/mm^2, 28.411 N/mm^2 and 27.661 N/mm^2 respectively falling short of the requisite compressive strength after 28 days.

***Keywords*—** Concrete mix design, Concrete strength, Fine aggregates.

I. Introduction

FINE aggregate material has been widely used for manufacture of concrete for use in buildings and other infrastructural developments. The acceptability of concrete as the most versatile product in construction is hinged on the availability of the respective material constituents, durability and the relative ease of its moulding to required shapes [1]. Concrete constitutes of Cement, fine and coarse aggregates and water. The aggregates form 75% of concrete by volume whose properties significantly affect the durability and structural performance of concrete [2]. The fine and coarse aggregate proportions vary depending on the design mix required for construction. Quality assurance of building materials is very essential in order to build strong, durable and cost effective structures [3]. Therefore the need to use right type and quality of aggregates in concrete manufacture cannot be underestimated; and the selection of the constituent materials should be made to the highest standard if the integrity of the structures is to be maintained [4].

The increased demand for housing and other infrastructural developments due to increase in population and urbanisation has resulted in high demand for aggregate for concrete production. Globally, material mined every year amounts to between 47-59 billion tonnes, with fine aggregate (sand) and coarse aggregate (gravel) accounting for the largest percentage (about 68- 85%), as well as the fastest increase in its exploitation rate [5]. River sand has been the most preferred choice of fine aggregate due to its availability, affordability and minimal or no processing requirements. A conservative estimate for world consumption of aggregates gives more than twice the amount of sediment carried by all of the rivers of the world [6], resulting in man being the planet's largest transforming agent with respect to aggregates [7]. This level of exploitation has led to increase in cost for concrete production and environmental degradation. The dredging of creeks, riverbeds and lake basins has resulted in ecological imbalance affecting bio- diversity and landscape, as well as having socio-economic, cultural and political consequences [8].

There is considerable pressure in many countries to use secondary and recycled aggregates in construction because of the environmental problems associated with production of primary aggregates (river sand) [9]. These include rock sand, quarry dust and manufactured sand. These fine aggregates are often manufactured by crushing and processing hard rocks to produce fine-grained materials. The degree to which the crushed rock sand can replace natural sand varies with rock type, the degree of quarrying processing used and the end use. [9]. In some quarries, the sand is washed to remove fines thereby significantly improving the quality.

Most developed countries use manufactured sand produced from crushing and processing of hard rock like limestone, sandstone and igneous rocks, whose aggregate properties are well researched. However due to the variance in geological processes that led to the formation of the parent rocks, the research findings cannot be applied to other areas because of the variation in rock mineral compositions. In Nairobi Metropolitan, the major sources of fine aggregates are Kajiado, Naivasha, Machakos, Mwingi and Mlolongo area. However, we have limited data on these material which could explain the poor quality concrete and structural failures [10]. This forms the

Kiambigi, MAINA, Department of Construction Management, JKUAT (phone: +2540722789750; e-mail: mkiambigi@ jkuat.ac.ke).

GWAYA, A.O, Department of Construction Management, JKUAT (e-mail: agwaya@ jkuat.ac.ke).

KOTENG, D.O, Technical University of Kenya.

basis of this research with a view to establishing the suitability of different aggregates used in Nairobi Metropolitan.

II. MATERIALS USED IN THE RESEARCH

The nature of concrete lies in the way the various materials are mixed, moulded and shaped to form a structure or element that can withstand its intended use over the lifetime of the structure without failure. Concrete volume is made up of 60-75% coarse and fine aggregates which significantly influence the fresh and hardened properties of concrete. The other components are Cement, water and additives (the latter is not considered in this study).

a) Cement

The cement used in the research was normal setting Ordinary Portland Cement of Class 42.5 strength, designated as OPC 42.5N. This is produced in accordance to KS EAS 18-1:2001, an adoption of the European Norm EN 197:2011 (Kenya Bureau of standards, 2005)

b) Coarse Aggregate

The coarse aggregate for the research was crushed aggregates of maximum size of 20mm sourced from Mlolongo quarry. Sieve analysis was done on the coarse aggregates in accordance to BS 812-1:1992. The percentage passing through the BS sieves was found to be within envelope limits. Coarse aggregates from the same source were used throughout the entire experiment.

c) Fine Aggregate

The fine aggregates used in the research were Natural river sand from Mwingi, Machakos and Kajiado, Quarry dust from Mlolongo, rock sand from Mlolongo and Naivasha sand from Naivasha quarry. The aggregates were graded in accordance to BS 812-1:1992. The Physical properties were determined in accordance to BS and ASTM standards; Bulk density (BS EN1097-3:1998), Specific gravity (ASTM C128), water absorption (BS 812-2:1995), Clay particles and friable materials (ASTM C142-97). The chemical properties were determined in accordance to BS EN 1744-1.

d) Water

Water used for concrete mixing and curing was obtained from Jomo Kenyatta university of Agriculture and technology treatment plant.

III. METHODOLOGY

The study employed experimental research divided into two parts; the first part was investigating the physical, chemical and mineralogical properties of fine aggregates collected from various sources; the second part involved determining the suitability of these fine aggregates for use in concrete. The design characteristic strength was maintained at 30N/mm^2.

a) Physical, chemical and mineralogical properties

The fine aggregates used in this research were identified as S2 (River sand obtained from Mwingi), S3 (River sand obtained from Kajiado), S4 (Rock sand obtained from Mlolongo), S5 (River sand obtained from Machakos), S6 (Naivasha Sand) and S7(Quarry dust obtained from Mlolongo). The physical properties were established in accordance with the British and American standards; while chemical properties were obtained using Atomic Absorption spectrometry (Varian Spectra AA10 machine) and validated using X-Ray Fluoresce method (Bruker S1 Titan machine) at the Ministry of Mining laboratory in Nairobi. The mineralogical properties were also determined using the X-Ray diffraction method (Bruker D2 Phaser machine) at the Ministry of Mining laboratory in Nairobi and counter checked with the geological formation of the catchment areas.

b) Concrete mix design

Concrete mix design using the different samples was done for class 30 concrete using D.O.E (Department of Environment)/British method at the JKUAT structures laboratory. This involved selecting and proportioning the constituents to give the required strength, workability and durability. [11]. Water/cement ratio, coarse aggregate/total aggregate ratio and total aggregate/cement ratio are the key parameters affecting design of a concrete mixture. For specified strength and durability requirements, a water/cement ratio has to be selected. [12] In this experiment a designed mix was used with strength testing forming an essential part of the requirements for compliance. [11] A characteristic strength of 30 N/mm^2 was specified with defective proportion of 2.5% yielding a standard deviation of 8N/mm2. A water /cement ratio of 0.52 was used (obtained from Table 2, Fig4 of the D.O.E) A slump of 10-30mm and a maximum crushed aggregate of 20mm was used yielding a free water content of 190 m^3 (Table 3 of the D.O.E). The aggregate was assumed to have a relative density of 2.7. The composition of Fine aggregate material was determined from the percentage passing Sieve no. 600µmm (Fig 6 – D.O.E). The respective constituents were then determined and varied based on the percentage of the material passing sieve no 600µmm.

Table 1 Concrete Mix Design of the samples

Sand type	Water/ Cement ratio	Water content (kgs)	Cement content (kgs)	Fine aggregate (kgs)	Coarse aggregates (kgs)
S2	0.52	190	365	656	1219
S3	0.52	190	365	525	1350
S4	0.52	190	365	788	1087
S5	0.52	190	365	656	1219
S6	0.52	190	365	562.5	1312
S7	0.52	190	365	787	1087

S2-Mwingi Sand, S3-Kajiado Sand, S4-Mlolongo Rock Sand, S5-Machakos Sand, S6-Naivasha Sand, S7-Mlolongo Quarry dust

IV. RESULTS AND DISCUSSION

a) Mineralogical composition of the aggregates

The mineral composition of the aggregate samples are shown in table 2.

Table 2 Mineral composition of fine Aggregates

Mineral	Formular	S2	S3	S4	S5	S6	S7
Labradorite	Al0.814 Ca0.32 Na0.18 O4 Si1.184	16.70%	16.40%		14.70%		
Berzalianite	Cu1.95 Se	25.40%	18.10%				
Berzalianite	Cu2 Se	23.40%					
Perryite	Fe0.24 Ni7.76 P0.63 Si2.37	5.20%					
Monipite	Mo Ni P	2.60%					
Quartz low	O2 Si	26.70%	21%		21.50%		
Quartz	O2 Si				20.40%		
Albite	Al Na O8 Si3		15.70%		10.20%		
Orligoclase	Al1.277 Ca0.277 Na0.723 O8 Si2.723		14.80%		11.90%		
Anorhtite Sodian	Al1.52 Ca0.52 Na0.48 O8 Si2.48		13.90%				
Polybasite	Ag31 As0.203 Cu S22 Sb3.797			2.60%			
Sanidine	Al Ba0.014 Fe0.003 K0.789 Na0.16 O8 Si3			14.60%			
Sanidine	Al K0.65 Na0.35 O8 Si3			13.30%			
Sanidine	Al1.04 Ca0.04 K0.65 Na0.31 O8 Si2.96					39.70%	28.10%
Sanidine	Al K O8 Si3						42.10%
Augite	Al Ca0.61 Fe0.13 K0.17 Mg0.43 Mn0.01 Na0.05 O6 Si1.61			6%			
Orthoclase	Al K O8 Si3			25.60%			
Microcline	Al K O8 Si3			17.30%			
Microcline	Al KO.95 NaO.05 O8 Si3				9.80%		
Diopside	Al0.078 Ca Fe0.024 Mg0.976 O6 Si1.922			5.80%			
Diopside	Ca Fe0.26 Mg0.74 O6 Si2					10.50%	
Augite	Al0.7 Ca Fe0.2 Mg0.6 O6 Si1.5			4.80%			
Bushmakinite	Al0.74 Cr0.26 Cu0.26 H O9 P1.22 Pb2 V0.52			1.10%			
Nepheline	Al3.84 K0.57 Na3.24 O16 Si4.16			7.50%			
Thorikosite	As0.2 Cl H0.5 O2 Pb1.5 Sb0.3			0.50%			
Smirnite	Bi2 O5 Te			1%			
Andesine	Al0.735 CaO.24 NaO.26 O4 Si1.265				11.40%		
Boulangerite	Pb10.159 S22 Sb7.841					5.20%	
Anorthoclase	Al K0.333 Na0.667 O8 Si3					46.50%	
Andorite VI	Ag Pb S6 Sb3						2.40%
Feldspar	Al1.9 O8 Si2.1 Sr						13.80%
Baricite	Mg3 O16 P2						13.60%

The aggregates mineral composition varied depending on the geology of the source. The mineral composition of the aggregates influences their chemical composition which in turn has an effect on the concrete strength development**.**

The mineral composition of S2 is comparable to the mineral composition of rocks at Mwingi catchment area whose geological formation indicates a highly metamorphosed series of sedimentary origin. The rocks mainly constitute quartz and iron oxides minerals [13].

The composition of S3 is similar to the mineral composition of the rocks at Kajiado catchment area whose geology indicates presence of volcanic rocks consisting of basalts, alkali trachytes and pyroclastics. These rocks are mainly composed of labradorite, orligoclase and quartz [14].

The composition of S4 and S7 is equivalent to the rocks at Mlolongo catchment area. The geological formation of the area indicates the presence of igneous and metamorphic rocks formed from volcanic eruptions. The rocks constitute mainly sanidine, orthoclase and microcline [15].

The mineral composition of S5 is analogous to the rock composition at Machakos- Mwala river catchment area. The geological formation of the area indicates presence of metamorphic series of pelitic, psammitic and calcareous rocks formed during the volcanic era. These rocks are mainly composed of diopside, biotite, dolomite, microcline and quartz. [15]

The mineral composition of S6 is equivalent to the rock composition of Naivasha (Suswa) catchment area. The geological formation of the area indicates presence of lavas, pyroclastic and lacustrine deposits formed during the volcanic era. The rocks are mainly composed of sanidine and boulangerite [16].

The geological formation of all the samples conform to the laboratory XRD results shown in Table 2, this indicates that true representative samples were obtained for testing in this experiment.

b) Physical composition of the aggregates

The physical characteristics of fine aggregates influences the properties of both freshly mixed and hardened concrete. The Fineness Modulus is a measure of the fineness of aggregates and is useful in determining the proportions of fine and coarse aggregates to be used in concrete mixtures. A higher fineness modulus implies a coarser aggregate hence requires more water to produce workable concrete [2].

Silt and clay content influences the strength development in concrete and should be maintained within the recommended limits. High silt and clay content has an effect on the resultant concrete as it will not achieve the expected strength [17].

Sieve analysis was used to determine the grading of the fine aggregates. The grading limits and the nominal maximum aggregate sizes are specified for fine aggregates as they affect the proportions of fine and coarse aggregates, water to cement ratio, workability and durability of the resulting concrete. Aggregates that do not have a large deficiency or excess of a particular size and give a smooth grading curve yields suitable concrete [18].

The bulk density of aggregate refers to the weight of aggregate divided by its volume. The volume includes that occupied by the aggregates as well as the voids between aggregate particles. The void content between the particles affect the paste requirements in the mix design. The specific gravity of aggregates refers to the ratio of its mass to the mass of an equal volume of water and is used in computations for mixture proportioning. [18]

The physical properties are shown in table 3.

Table 3 Physical properties of fine Aggregates

Sno.	Test Parameter	S-2	S-3	S-4	S-5	S-6	S-7
1	Specific gravity	2.12	2.06	2.24	2.31	**1.73**	2.27
2	Apparent Specific gravity	2.57	2.5	2.6	2.63	2.36	2.59
3	Bulk density	1497	1469	1407	**1613**	1327	**1684**
4	Water Absorption	8.3	8.62	6.31	5.16	15.3	5.37
5	Fineness Modulus	2.66	1.92	3.37	2.54	1.94	3.66
6	Silt and clay content	4.85	4.16	2.06	6.66	9.37	11.9
7	Sieve Analysis	C&M	F	C	C&M	F&M	C
8	Surface texture	Rough	Smooth	Coarse	Rough	Smooth	Coarse
9	Particle shape	R	R	A	R	R	Fl & E

C- Coarse, F-fine, M-medium, R-rough, A-angular, Fl-flaky, E-elongated

The specific gravity of all the aggregates samples fell below the minimum accepted limit of 2.6 required in a concrete mix according to the ASTM C127 standards. The water absorption of the all aggregate samples exceed the 2.3% maximum accepted limit according to ASTM C127 standards. The highest was recorded in S6 while S5 had the least. The higher the water

absorption the higher the amount of water required to produce workable concrete.

The silt and clay content in all the aggregate samples exceeded the maximum limit of clay content according to the ASTM C142 standard that stipulate a limit of 1% maximum. The highest concentration was recorded in S7 while the lowest was in S4. The high silt and clay content in S7 reduces concrete strength.

The fineness modulus of all the aggregate samples ranged between 1.93- 3.66. The standard limits of fineness modulus are 2.3-3.1 according to ASTM C33.The highest Fineness Modulus was recorded in S7 indicating that the aggregate is coarse and will require more water to produce workable concrete. The lowest fineness modulus was recorded in S3 therefore will require less water.

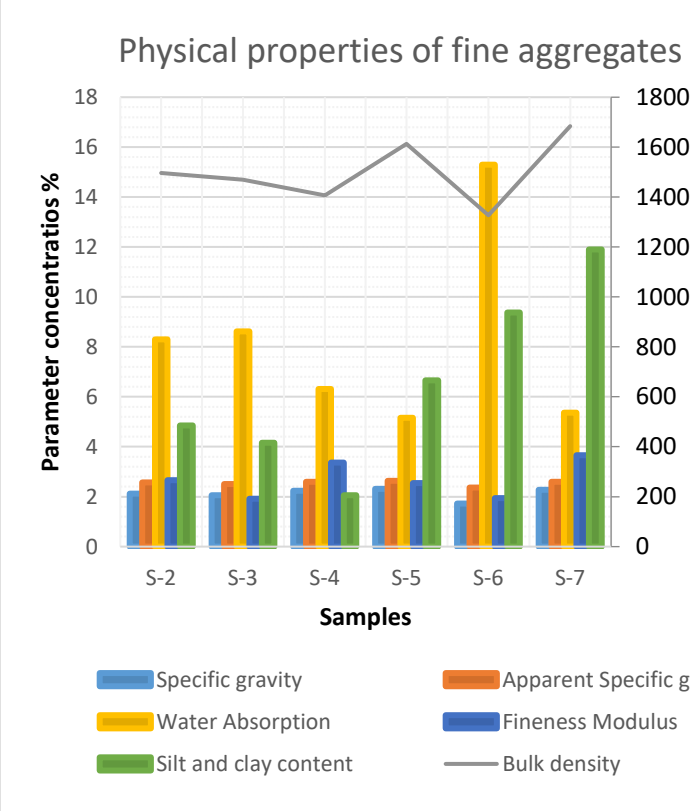


Figure 1Physical Properties of Fine Aggregate

The sieve analysis of the fine aggregates were as shown in figure 2. It was observed that the samples fell within the limits as per the requirement of BS 812-1:1992 for natural aggregates.

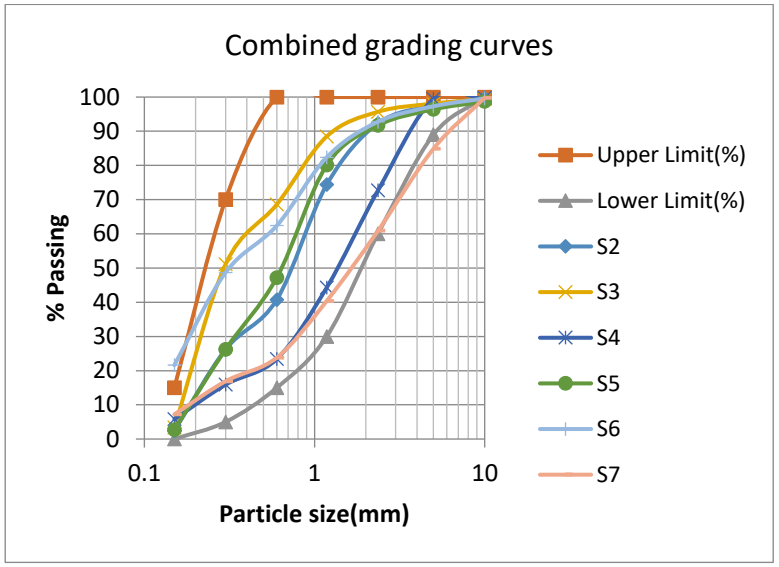


Figure 2 Grading Curves of Fine Aggregates

c) Chemical Properties of Aggregates

The chemical properties of aggregates have a great influence on strength development of concrete. The main chemical constituents are Silicon IV oxide, Aluminum III oxide and calcium oxide which influence the setting time, early strength and final concrete strength. Silica concentrations of between 70-90% prolongs the setting time but increases the final concrete mix strength [19], alumina concentrations of between 8-12% reduce the setting time but increases the concrete strength [19] and Calcium oxide concentrations of between 2-5% prolong the setting time of concrete but gives an early strength. [19].

The chemical properties of the fine aggregates are shown in table 4.

Table 4 Chemical Properties of Fine Aggregates

Sno.	Parameter (%)	S-2	S-3	S-4	S-5	S-6	S-7
1	Sio_2	76.00	78.00	67.00	80.00	69.00	65.00
2	Al_2O_3	11.00	9.00	17.00	10.00	14.00	19.00
3	Fe_2O_3	1.40	1.20	4.00	1.00	5.50	4.00
4	CaO	1.60	1.50	1.40	2.50	1.30	1.40
5	MgO	0.80	1.00	0.05	0.02	0.04	0.08
6	Na_2O	2.00	1.40	1.50	1.80	3.00	4.00
7	K_2O	1.00	1.00	3.00	1.00	1.80	1.60
8	TiO_2	0.30	0.17	1.40	0.12	0.30	0.60
9	Lol	0.72	1.04	3.50	1.70	2.00	3.80

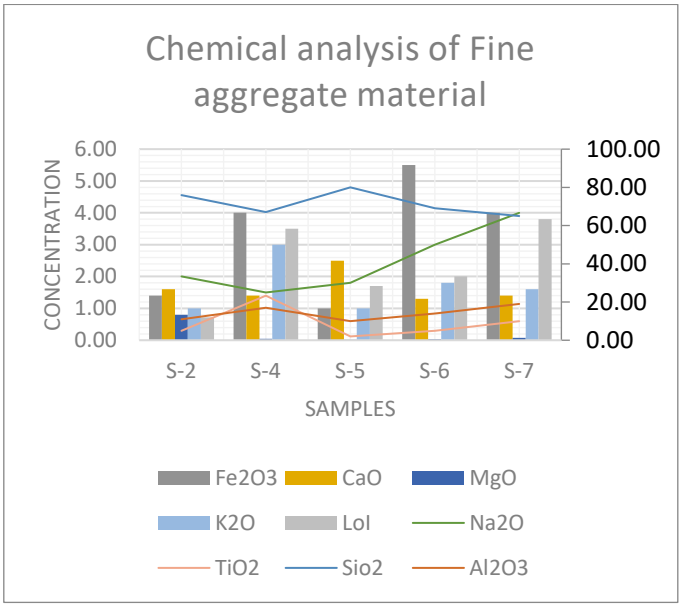


Figure 3 Chemical Properties of Fine Aggregates

The Silicon IV oxide concentration in the aggregate samples varied between 65-80%. The adequate Silicon IV oxide concentration for fine aggregates in concrete production is 70-90%. The highest concentration (80%) was recorded in S5 and it is expected to develop a high final concrete strength. The lowest concentration was recorded in S7 (65%) which would result in reduced final concrete strength.

The Aluminium III oxide concentration varied between 9-19%. For concrete production, the adequate Aluminium III oxide concentrations in fine aggregates is 8-12%. The highest was recorded in S7 (19%).The lowest was recorded in S3 (9%).

The calcium oxide concentration for all aggregate samples varied between 1.3-2.5%. The adequate calcium oxide concentrations in fine aggregates for concrete production is 2-5%. The highest was recorded in S5 (2.5%) and it would therefore develop a higher early strength (7days) .The lowest was recorded in S6 (1.3%) and it would therefore have reduced early concrete strength.

The variation in chemical properties in the various aggregate samples can be attributed to the mineral compositions of the parent rock.

d) Compressive strength of concrete from the samples.

The compressive strength developed by the concrete produced from different fine aggregates samples is shown in table 5.

Table 5 Compressive Strength development with curing age

Days	S2	S3	S4	S5	S6	S7	characteristic target strength
7	24.756	25.395	20.893	21.78	20.89	20.032	20
14	32.146	31.196	25.121	27.617	24.748	26.064	27
28	41.899	37.173	28.682	33.645	28.411	27.661	30

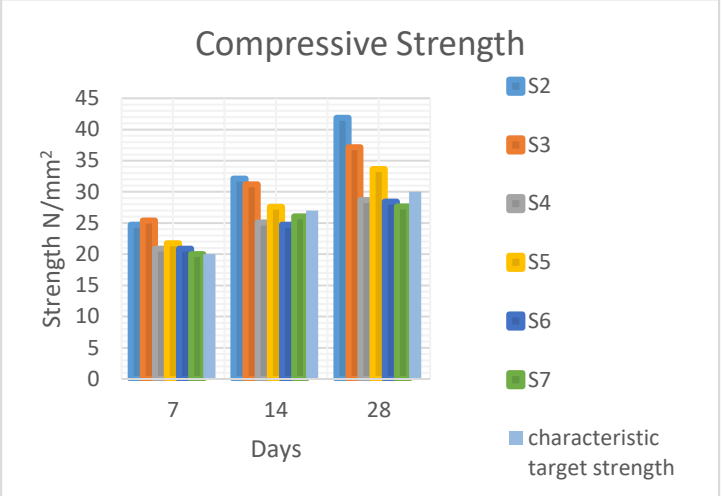


Figure 4 Compressive Strength development with curing age

The compressive strength test was done at 7, 14 and 28 days to assess strength development of concrete. All samples indicated a progressive strength gain with curing age with Sample S3 developing the highest initial strength and S7 recording the lowest. At 7 days curing, all the samples attained early strength of 65% (20N/mm2) of the characteristic strength which can be attributed to the presence of calcium oxide in all the samples thereby contributing to early strength. The hot climate of the area also promoted early strength [2].

At 28 days curing, samples S2, S3 and S5 obtained 100% of the characteristic strength (30N/mm2) while samples S4, S6 and S7 did not achieve the characteristic strength. Samples S2, S3 and S5 had quartz as the predominant mineral which could contributed to high concentration of Silicon IV oxide. Though this prolongs the setting time of concrete it eventually increases the final concrete mix strength. Samples S4, S6 and S7 had lower concentrations of Silicon IV oxide, which contributes to the lower strength in 28 days.

The choice of aggregates used influenced compressive strength development in the concrete samples. The compressive strength results indicate that concrete strength is influenced by the chemical composition of the aggregates, grading and the particle shape and texture.

V. CONCLUSIONS

The compressive strength and other properties of concrete are influenced by the physical, chemical and mineralogical properties of the fine aggregates used, the mix design, curing and the concrete placement method. Observations made during this research indicate that different fine aggregates vary in chemical, physical and mineralogical properties depending on the area of source and the weathering and crushing processes. The observations made on the resultant concrete indicate that concrete manufactured using natural river sand obtained higher compressive strengths than the concrete manufactured using alternative aggregates when the slump and water/cement ratio are constant at 28 days curing. All natural river sand sample aggregates obtained from Mwingi, Machakos and Kajiado area attained the target strength at 28 days and therefore suitable for use in the construction industry. Fine aggregates obtained from

Naivasha, quarry dust and rock sand obtained from Mlolongo did not meet the design target strength requirement at 28days with an average deviation 5-6%. This can be explained by the inherent physical and chemical properties of these aggregates especially due to low Silica concentration, high silt and fineness modulus.

REFERENCES

[1] Civil engineers forum, "Advantages and disadvantages of Concrete,"2016.[Online].Available:http://civilengineersforum.com/concrete-advantages-disadvantages/. [Accessed 15 February 2018].

[2] A. M. Neville, PROPERTIES OF CONCRETE, LONDON: PEARSON EDUCATION LIMITED, 2011.

[3] A. Savitha, "Importance of quality Assurance of Materials for Construction work," Building materials Research and Testing division, 2012.

[4] The Constructor, "Manufactured Sand (M-Sand) in Construction," 2016.[Online].Available:http://theconstructor.org/building/manufactured-sand-m-sand-in-construction/8601/. [Accessed 22 March 2016].

[5] F. Krausman, S. Gingrich, N. EisenMenger, K. H. Erb, H. Haber and M. Fischer-Kowalski, "Growth in Global Material use, GDP and Population During the 20th Century," Ecological Economics, vol. 63, pp. 2696-2705, 2009.

[6] D. Milliman and M. Syvitski, "Geomorphic/ Tectonic Control of Sediment Discharge to the Ocean: The Importance of Small Mountainous Rivers," J. Geol., no. 5, pp. 525-544, 1992.

[7] R. T., "Special Edition for the World Conference on Disaster Reduction: It's the way you tell them, Environment & Poverty Times," Kobe, Japan, Jan 2005.

[8] S. Sonak, P. Pangam, M. Sonak and D. Mayekar, "Impact of sand Mining on Local Ecology in Multiple Dimension of Global Environmental Change," The Energy and Resources Institute, New Delhi, India, 2006.

[9] E. s. DJ Harrison, "Alternative sources of Aggregates," British geological survey, Nottingham, 2003.

[10] Ministry of Nairobi Metropolitan, "Safety and Security of the Built Environment," GoK, Nairobi, 2013.

[11] B. C. T.J MacGINLEY, Reinforced Concrete Design Theory and Examples, London: E & FN Spon, 2003.

[12] S. Ahmad, "Optimum Concrete mixture Design using locally available ingredients," The Arabian jounal of Science and Engineers, 2007.

[13] A. F. Crowther, "Geology of the Mwingi Area," Geological Survey of Kenya, Nairobi, 1957.

[14] F. J. Matherson, "Geology of Kajiado Area," Geological Survey of Kenya, Nairobi, 1964.

[15] D. L. Searle, "Geology of thee Sultan Hamud Area," Geological Survey of Kenya, Nairobi, 1953.

[16] R. D. A.O. Thomsons, "Geology of the Naivasha Area," Geological survey of kenya, Nairobi, 1958.

[17] V. A. Okumu, "The Effects of the Properties of Constituents Materials on the Quality of Concrete in Kenya," in Annual Conference on Sustainable Research and Innovation, Nairobi, 2016.

[18] University of Memphis, "Aggregates for Concrete," [Online]. Available:www.ce.memphis.edu/1101/notes/concrete/PCA_manual/Chap05.pdf. [Accessed 15 February 2018].

[19] Penn State University, "The Effect of Aggregate Properties on Concrete,"2000.[Online].Available:http://www.engr.psu.edu/ce/courses/ce584/concrete/library/materials/aggregate/aggregatesmain.htm.

[20] O. Olanitori.L M, "The Effect of Clay Impurities in Sand on the Crushing Strength of Concrete,(A Case Study of Sand in Akure metropolis, Ondo state, Nigeria)," in 30th Conference on Our world in Concrete and Structures, Singapore, 2005.

A Review on Prognosis of Rolling Element Bearings Operated Under Non-stationary Conditions

Leila L. Mbagaya, James K. Kimotho and Jackson G. Njiri

Abstract—Bearings constitute a majority of the components found in rotating machines. Though inexpensive, their failure can result in unnecessary downtime, losses in production and propagation of failure to other critical components leading to expensive maintenance actions. Most of these machinery are operated under adverse and varying conditions which result in difficulty in defining health indices from condition monitoring data. Therefore, techniques for condition monitoring of rotating machinery operated under non-stationary conditions are necessary. This paper aims at summarizing and reviewing the recent trends and developments in prognostics of rotating machines operating under non-stationary conditions with emphasis on rolling element bearings. Various techniques, methods, and models used in bearing prognosis are discussed. Moreover, the research gaps and possible future trends are addressed in the conclusion.

Keywords—Bearing Prognosis, Non-stationary Conditions, Remaining Useful Life (RUL),

I. INTRODUCTION

ROTATING machineries are the most common mechanical components in industrial application. Their main components are gear boxes, roller bearings and rotary shafts. Typically, these machines operate under adverse conditions of high load and high temperatures. These conditions may cause severe breakdowns and decrease in the equipment's performance resulting in reduced safety, availability and reliability, economic losses, and lower product quality [1]. A good maintenance strategy is crucial in preventing such catastrophic failures while maintaining machine safety and reliability as well as adding value to maintenance practices.

The simplest maintenance strategy employed in the industries is known as breakdown maintenance according to Heng et al [2]. This is where machines are run until they fail and when failure has occurred, reactive maintenance is carried out. This approach can be extremely costly due to long hours of machine downtime and may also lead to propagation of failure to other components. A slightly more effective time-based maintenance technique known as preventive maintenance involves periodical cleaning, servicing and inspection of machines in order to prevent abrupt failure. However, this method cannot guarantee the absence of any breakdown and the replacement of parts before their end lifetime.

L. L. Mbagaya, Department of Mechatronic Engineering, JKUAT (e-mail: leila.mbagaya@jkuat.ac.ke).
J. K. Kimotho, Department of Mechanical Engineering, JKUAT (e-mail: jkuria@eng. jkuat.ac.ke).
J. G. Njiri, Department of Mechatronic Engineering, JKUAT (email:jackgithu@eng.jkuat.ac.ke).

More efficient maintenance approaches such as condition-based maintenance (CBM) have been adopted to address the issue of machine reliability and reduction of maintenance related expenses. CBM is a maintenance strategy aimed at maximizing productivity and machine up time while lowering operating costs by carrying out maintenance when the need arises [3]. The actual conditions of a machine are monitored to obtain the health status of a system and if the indicators show signs of upcoming machine failure, maintenance is carried out. A CBM program consists of two important aspects known as diagnostics and prognostics.

Diagnostics is concerned with fault detection, isolation and identification when it occurs. Fault detection indicates when a failure has occurred; fault isolation locates the faulty component; and fault identification determines the nature of the fault when it is detected. Prognostics, unlike diagnostics, is a prior event analysis process that deals with failure prediction before it occurs. It is the ability to predict accurately and within acceptable error bandwidth the remaining useful life of a failing component or subsystem [4]. Prognostics is much more efficient than diagnostics in reducing machine downtime. However, diagnostics is still useful when failure prediction fails and a fault occurs.

Prognostic approaches can be classified into three main methods as shown in Fig. 1. Data-driven approach relies on

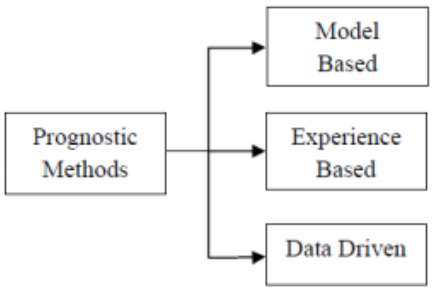


Fig. 1: Classification of Prognostic Methods [5].

observable past data and statistical models [6]–[8]. The models are derived from routinely monitored system operating data such as oil debris, vibration signals, temperature, and pressure.

Experience based approach is based on the use of simple reliability functions such as Weibull law and exponential law, rather than complex mathematical models [9]. Model-based approach however, uses models which make use of laws of physics [10]–[12]. This approach employs residuals as features by carrying out consistency checks between sensor measurements of a system and outputs of a mathematical model.

Most of the research that has been carried out in the field of prognostics has focused on machines operating under stationary conditions. Rotating machines have increased complexities and complex degradation processes due to non-stationarity. Satisfactory results can therefore not be produced when traditional techniques are applied to non-stationary conditions. With increasing complex machinery, there arises a need for CBM techniques that are able to operate on such machines operating under non-stationary conditions such as wind turbines and automobile drive trains. Thus the focus on this paper is to review the prognostics of rotating machines operated under non-sationary conditions. The key subject of this study is rolling element bearings because they have resulted in a majority of the failures in rotating machines [13] as depicted in Fig 2. Moreover, failure prediction of bearings can improve the safety and reliability of rotating machines while reducing maintenance costs.

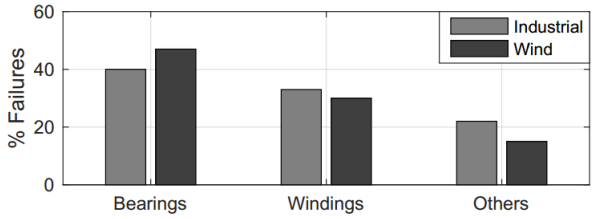


Fig. 2: Comparison of failures in rotating machines in industrial and wind turbine applications [14]

II. PROGNOSTIC APPROACHES

A. Model-based Approach

The model-based prognosis approach relies on a mathematical model of system under observation by assuming that a fault in the system will lead to deterministic changes in the model parameters. Input includes information on operating and environmental conditions. A comparison of model output to actual system outputs is done to generate a residual signal as depicted in Fig. 3. The ratio of output and input can be used as a health index to track degradation of the system. Based on that generated signal, useful information is extracted and potential fault conditions are identified.

A common model-based approach is crack growth modelling. Li et al [16] based their bearing prognostic methodology on the in process adaptation of defect propagation rate with vibration signal analysis. The defect size as predicted by

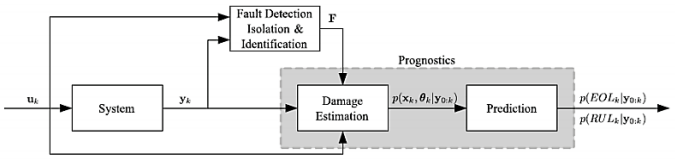


Fig. 3: Model-based prognostic approach [15].

a fatigue crack propagation model was compared to the estimation from a diagnostic model in the future to fine tune the propagation model parameters. However, the assumption that the defect size can directly be estimated from vibrations is faulty since the instantaneous defect size cannot be measured without interrupting machine operation.

Li et al [17] predicted spall progression of tapered roller bearings using an empirical method. The empirical constants need to be determined for all bearings and all operating conditions for which is used. Kotzalas and Harris [18] also presented a spall progression model by extending the Ionnades-Harris (I-H) fatigue life theory. The equations relating spall progression rate $\frac{dSp}{dN}$ to spall similitude W_{sp} are as follows;

$$\frac{dSp}{dN} = C(W_{sp})^m \tag{1}$$

$$W_{sp} = (\theta_{max} + \tau_{avg})\sqrt{\pi Sp} \tag{2}$$

where θ_{max} is maximum stress, τ_{avg} is average shearing stress, Sp is spall length, C and m are constants.

The research done by Kotzalas and Harris [18] showed that 3 to 20% of a particular bearings useful life remains after spall initiation. The study identified two spall progression regions as shown in Fig. 4; stable spall progression region characterized by gradual spall growth and minimum vibratory loading and unstable spall region characterized by increasing broadband vibration amplitudes. The boundary between the two regions was selected as the spall progression life. With their model, prediction of the life of a spall progression was achieved so long as bearing fracture does not occur.

Li and Lee [11] used an embedded gear dynamic model to predict the remaining useful life (RUL). The advantage of this model is that finite element analysis (FEA) enables stress calculation based on the gear geometry, speed, load, material properties and so on. However, this method is time consuming, needs expensive software to analyze the vibration data and calculate the stress value, and the results rely on the accuracy of the defect size.

Oppenheimer and Loparo [19] used Forman Law of linear elastic fracture mechanics to model rotor shaft crack growth. The assumption made from these crack growth models is that the defect size could be estimated directly from vibration data [2]. This assumption, however, is questionable

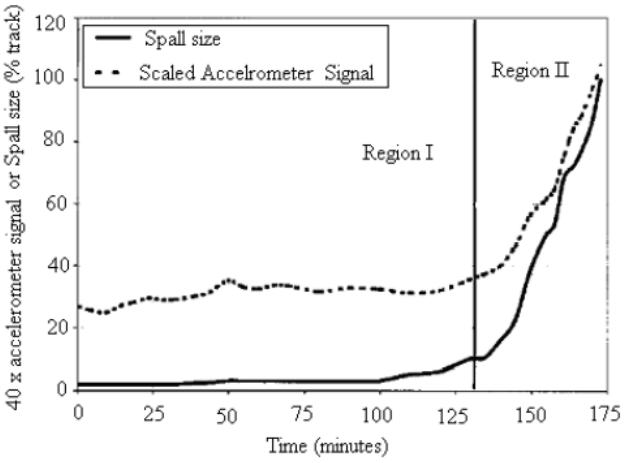


Fig. 4: Scaled accelerometer signal and spall size as a function of time

since instantaneous defect size cannot be measured without interrupting the operation of the machine thereby hindering the usefulness of this model.

Linkan Bian et al [20] investigated a method for modeling degradation signals from components functioning under dynamically-evolving environment conditions. In-situ sensor signals were utilized in real time to predict and update the distribution of a component's residual lifetime. The research showed that the Bayesian updating scheme provides reasonable lifetime prediction results, especially as information is progressively revealed over time. However, this is based on the simplified assumption that the current environmental or operational conditions affect the time-dependent rate at which a component's degradation signal increases.

Matej Gasperin and Juricic [21] modelled feature time series as an output of dynamic state model. The model was then used to determine the presence of a fault and predict the future behavior and remaining useful life of a system. The optimal model at the current state of failure is found by adopting an algorithm for on-line model estimation. The approach is validated using the experimental data on a single stage gearbox. The results showed that the model can be used to predict the evolution of the fault under variable operating conditions, if the future time profile of the load is known. Moreover, a linear relationship was assumed between operating conditions, fault dimension and vibration feature value.

The research done by Liao and Tian [22] was also based on simplified assumptions on the relationship between operating conditions and the rate of degradation. An enhanced Bayesian technique for predicting the RUL of a single unit under time-varying operating conditions was investigated. The approach integrates in situ degradation measurements of the interested unit as well as the operating conditions with a population-based Accelerated Degradation Testing (ADT) model. The results showed that the proposed approach is capable of achieving accurate RUL prediction under complex operating conditions that may involve stochastic components. However, more test units need to be considered and further investigation into the different failure modes needs to be done.

Zhao et al [23] developed an integrated prognostics approach to deal with time-varying operating condition, which integrates physical gear models and sensor data. The degradation model is built on the physics of damage progression, which takes the form of a function of environmental parameters. Any changes of these environmental parameters, such as load, temperature, and speed can be manifested immediately in the physical model. The assumption that future loading conditions are known may lead to difficulty in quantifying the remaining useful life. Moreover, validation of the proposed model with experimental investigations in a lab environment did not take place. Therefore this model may not represent the physical behavior of the target system.

B. Experience-based Approach

Experience-based prognostic approach is concerned with integrating reliability data into prognostics. It uses the data of the experience feedback gathered during a significant period of time (maintenance data, operating data, failure times, etc.) to adjust the parameters of some predefined reliability models. The obtained models are then used to predict the time to failure, or the RUL cite.

Goode et al [24] presented a relatively simple model that combines reliability data with condition monitoring measurements to predict the remaining useful life of pumps in a hot strip steel mill. The evolution of CM data was divided in stable and failure zone. Little information was provided by CM data in stable zone. Therefore reliability data was used to predict the point when machine enters failure zone. The results from the study indicated that the prediction model is dependent on the quality and accuracy of the condition-monitored measurements.

Proportional Hazard (PH) models have also been employed in prognostics by Jardine et al [25], [26] to predict the reliability of rolling element bearings and engines. PH models are a class of survival models in statistics where analysis deals with time duration until one or more events happen. These models assume that hazard changes proportionately with asset condition and that the proportionality constant is the same at all time.

Wang and Christer [27] modeled bearing residual life distribution based on stochastic filtering theory. The model developed in the paper was based upon the assumption that the measurement noises are non-Gaussian distributed, which is a natural requirement in condition monitoring modelling. However, as

with all other cases employing condition monitoring (CM) and reliability data, threshold identification of defect initiation is required, which is challenging to determine.

C. Data-driven Approach

Data-driven methods are based upon statistical and learning techniques and are derived from routinely monitored system operating data such as oil debris, vibration signals, temperature, and pressure. Most of the data-driven approaches originated from the theory of pattern recognition [3]. They mainly comprise of Artificial intelligence (AI) techniques and statistical methods [1]. Statistical methods include state space models (e.g Bayesian networks [28], hidden Markov Models (HMM) and hidden semi-Markov Models (HSMM) [29]) and regressive models while AI techniques include neural networks [30].

Figure 5 illustrates the stages in a date-driven approach. Data acquisition involves measuring the appropriate form of data. The measured condition monitoring data can be vibration data, acoustic data, oil analysis data, etc. The data measured is polluted by different types of noise. Pre-processing removes the noise through filtering and prepares it for feature extraction. Feature extraction involves processing the filtered data. It can be performed in the time-domain, frequency-domain or time-frequency domain. After this has been done, post-processing is then carried out to prepare feature vectors for pattern-recognition stage. Pattern recognition is where a method is applied to decide the damage state based on the feature vectors extracted by signal processing techniques.

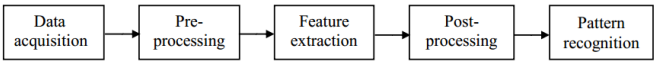


Fig. 5: Data-driven prognostic approach

Data-driven approach is advantageous over model-based approach in cases where the system is complex and thus accurate modeling becomes expensive. Moreover, data-driven approach is applicable where an understanding of first principles of system operation is not comprehensive. However, the primary drawback of such an approach is that effectiveness is not only extremely dependent on the quantity but also quality of system operational data. The systems require large amount of training data and it may have wider confidence intervals in comparison to other approaches. Furthermore, it is difficult to obtain run-to-failure data particularly for new systems because running systems to failure could be lengthy and costly.

Gebraeel et al [6] predicted bearing failure time by using the ANN approach. An experimental setup was developed to perform accelerated bearing tests where vibration information was collected from a number of bearings that are run until failure. This information was then used to train neural network models on predicting bearing operating times. Vibration data from a set of validation bearings were then applied to these network models. The resulting predictions were used to estimate the bearing failure time. Comparisons between these predictions with the actual lives of the validation bearings and errors were computed to evaluate the effectiveness of each model. The results showed that 64% of the predictions were within 10% of actual bearing life, while 92% of predictions were within 20% of the actual life. However, the drawback of this method is that the failure thresholds were not adequately defined.

A trained dynamic wavelet neural network (DWNN) was employed by Vachtsevanos and Wang [31] in prognosis of a defective bearing with a crack in its inner race. It was noted that more extensive failure data, that is difficult to obtain in critical processes, is required to draw firm and comparative conclusions.

A Recursive bayesian technique was proposed by Zhang et al [32] to estimate asset health reliability using condition monitoring data. This method enabled reliability evaluation using observations from individual assets, rather than using failure data from a population of assets. Validation of the employed method was implemented by an experiment on bearing life testing as a case study. The accuracy of such a technique relies strongly on the correct determination of thresholds for various trending features as depicted in [2].

Hidden Markov models (HMMs) were integrated with an adaptive stochastic fault prediction model and principal component analysis (PCA) and used in bearing prognosis by Zhang et al [33]. The principal features extracted by PCA were utilized by HMM to generate a health/degradation index representing the current system health status. This was then used as an input to an adaptive prognostics component for on-line remaining useful life prediction. The merit of this approach is the on-line learning capability which increases its prediction accuracy. However, the inability to physically observe a defect n an operating unit makes it difficult to relate the defined health-state change point to the actual defect progression. Chinam and Baruah [34] also employed HMMs to model degradations on bearings, and to estimate the underlying RUL.

Dong and He [29] presented a statistical modelling methodology based on segmental hidden semi-Markov models (HSMMs). An HSMM is a hidden Markov model (HMM) with temporal structures. However, unlike HMMs, HSMMs employ explicit probability distributions such as Gaussian distribution to model the state durations more accurately. The developed method was then tested using data from a real hydraulic pump health montoring application case study. The results showed that the recognition rates for all states were greater than 96%. For each individual pump, the recognition rate increased by 29.3% in comparison with HMMs. However, this method relies on the assumption of predetermined failure threshold in order to carry out prognostics.

III. PROGNOSTIC TECHNIQUES

The various techniques that have been successfully employed for prognostics include vibration analysis, oil analysis, temperature analysis, acoustic emissions and so on. These methods are effective in describing machine performance.

A. Vibration-based Signal Processing

Vibration analysis has been used to predict the RUL of bearing by use of current and previous vibration data and for diagnosis of all types of fault, either localized or distributed. Vibration-based signal analysis can be performed in the time domain, the frequency domain or the time-frequency domain.

Monitoring the variation in statistical indices such as kurtosis, root mean square (RMS) value or crest factor can help detection of bearing faults in the time-domain analysis [35]. The disadvantage with this method is that it is difficult to determine appropriate thresholds which should not be exceeded because variations exist in different applications. Frequency-domain analysis is based on time-frequency transformation and the most popular diagnostic method uses Fourier Transform. The presence of fault characteristic frequency indicates a fault in bearing diagnosis. The main disadvantage of frequency domain analysis is inability to locate particular frequency in time domain. To overcome this problem, time-frequency analysis is used.

Time-frequency domain techniques can be used to analyze non-stationary signals. A popular time-frequency analysis is the Short Time Fourier Transform (STFT) [36] which is a Fourier-related transform that determines the sinusoidal frequency and phase content of local sections of a signal as it changes over time. However, this method is limited in its time-frequency resolution. Cocconcelli et al [37] enhanced the fault signature of a ball bearing under varying motor-speed by averaging the short-time fourier transform (STFT) for each shaft revolution in the time-frequency domain. The sum of the averaged STFT coefficients was used as an indicator of the level of damage on the bearing. However, the relationship between the damage indicator and varying shaft speed is lacking.

Feng and Liang [38] presented a time-frequency analysis method based on the Vold-Kalman filter and higher order energy separation (HOES) to extract fault symptoms in a wind-turbine gearbox under non-stationary conditions. The results showed that it was effective in diagnosing gear faults. However, investigation on how the faults evolve with time was not done.

An alternative to STFT is the Wavelet Transform (WT) [39] which has more flexible time-frequency resolution and is more applicable in fault detection. This method can be classified according to signal decomposition paradigms as continuous WT, discrete WT and wave packet analysis. Gritli et al [40] proposed the use of discrete wavelet transform (DWT) with high multiresolution analysis (HMRA) of stator currents for fault diagnosis of rotors in doubly fed induction machine (DFIM). However, this approach was only evaluated for fault diagnosis.

Guan et al [41] presented a time-frequency method that outperforms others in providing fine-resolution time-frequency preparation. The synchrosqueezing transform-based method was effective in detecting distributed and localized gear faults under nonstationary conditions. However, the method was not evaluated on ability to track evolution of the faults.

Antoniadou et al [42] presented a time-frequency analysis approach for condition monitoring of wind turbine gearboxes under varying operating condition. The Empirical Mode Decomposition (EMD) method was used to decompose the vibration signals into meaningful signal components associated with specific frequency bands of the original signal. Furthermore, the Teager-Kaiser energy operator (TKEO) approach was employed to improve the estimation of instantaneous spectral characteristics of the vibration data under certain conditions. In this approach, the relationship between the operating conditions and the features is assumed.

B. Oil Analysis

Wang and Zhang [43] predicted the residual life of aircraft engines monitored based upon available oil monitoring information. The fundamental concept behind the model is the proportional residual life that assumes the residual life is proportional to the actual wear measured by the oil analysis programmes. The oil analysis data used was the total metal concentration obtained using Spectrometric Oil Analysis Programme (SOAP) from aircraft engines. The principal component analysis PCA was applied to preanalyze the data. The goodness-of-fit test was then carried out to test the model established. The results obtained from the analysis showed that it is feasible to model the relationship between residual life and information obtained from an oil analysis program. However, this model required the determination of a threshold level to indicate defect initiation point, which is in practice, difficult to determine.

Orchard and Vachtsevanos [9] employed particle filtering for prognosis in turbo engine. Particle filtering was used in the proposed method as CM data to monitor turbine blade health. The particle filtering algorithm consecutively updated the current state estimate for a nonlinear state-space model (with unknown time-varying parameters), and predicted the evolution in time of the probability distribution for the crack length. Authors reported acceptable results in terms of precision and accuracy. However, this method registers poor performance with high dimensional data.

IV. RESEARCH GAPS

1) The research does not depict how the faults evolve with time and failure thresholds under non-stationary conditions are not defined.

2) Some of the research work done employs simplified assumptions on the relation between operating conditions and the rate of degradation
3) Some of the research work with proposed models under non-stationary conditions has not been validated with experimental investigations.

V. CONCLUSION

This paper has briefly reviewed the progress in the research and development of rotating machinery prognostics and especially rolling element bearings. The review indicates that most of the existing prognostics studies have mainly concentrated on systems operated under stationary conditions. However, in real-life situations, most systems operate under non-stationary conditions. The effects of variations in operating conditions have not been properly explored and modelled for fault prognosis. Further research is also required particularly in cases where the faults evolve with time and failure thresholds are not defined. Moreover, the models discussed above need validation with experimental investigations in order to verify their application to industrial problems.

ACKNOWLEDGEMENT

We acknowledge the staff and management of Jomo Kenyatta University of Agriculture and Technology for their support by allowing the authors to use their research facilities and resources to accomplish this review paper.

REFERENCES

[1] A. K. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical systems and signal processing*, vol. 20, no. 7, pp. 1483–1510, 2006.

[2] A. Heng, S. Zhang, A. C. Tan, and J. Mathew, "Rotating machinery prognostics: State of the art, challenges and opportunities," *Mechanical systems and signal processing*, vol. 23, no. 3, pp. 724–739, 2009.

[3] Y. Peng, M. Dong, and M. J. Zuo, "Current status of machine prognostics in condition-based maintenance: a review," *The International Journal of Advanced Manufacturing Technology*, vol. 50, no. 1, pp. 297–313, 2010.

[4] J. Sikorska, M. Hodkiewicz, and L. Ma, "Prognostic modelling options for remaining useful life estimation by industry," *Mechanical Systems and Signal Processing*, vol. 25, no. 5, pp. 1803–1836, 2011.

[5] D. A. Tobon-Mejia, K. Medjaher, N. Zerhouni, and G. Tripot, "A mixture of gaussians hidden markov model for failure diagnostic and prognostic," in *Automation Science and Engineering (CASE), 2010 IEEE Conference on*, pp. 338–343, IEEE, 2010.

[6] N. Gebraeel, M. Lawley, R. Liu, and V. Parmeshwaran, "Residual life predictions from vibration-based degradation signals: a neural network approach," *IEEE Transactions on industrial electronics*, vol. 51, no. 3, pp. 694–700, 2004.

[7] N. Z. Gebraeel and M. A. Lawley, "A neural network degradation model for computing and updating residual life distributions," *IEEE Transactions on Automation Science and Engineering*, vol. 5, no. 1, pp. 154–163, 2008.

[8] N. Z. Gebraeel, M. A. Lawley, R. Li, and J. K. Ryan, "Residual-life distributions from component degradation signals: A bayesian approach," *IEEE Transactions*, vol. 37, no. 6, pp. 543–557, 2005.

[9] M. E. Orchard and G. J. Vachtsevanos, "A particle-filtering approach for on-line fault diagnosis and failure prognosis," *Transactions of the Institute of Measurement and Control*, vol. 31, no. 3-4, pp. 221–246, 2009.

[10] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems. reviews, methodology and applications," *Mechanical systems and signal processing*, vol. 42, no. 1, pp. 314–334, 2014.

[11] C. J. Li and H. Lee, "Gear fatigue crack prognosis using embedded model, gear dynamic model and fracture mechanics," *Mechanical systems and signal processing*, vol. 19, no. 4, pp. 836–846, 2005.

[12] G. Kacprzynski, A. Sarlashkar, M. Roemer, A. Hess, and B. Hardman, "Predicting remaining life by fusing the physics of failure modeling with diagnostics," *JOM Journal of the Minerals, Metals and Materials Society*, vol. 56, no. 3, pp. 29–35, 2004.

[13] A. H. Bonnett and C. Yung, "Increased efficiency versus increased reliability," *IEEE Industry Applications Magazine*, vol. 14, no. 1, 2008.

[14] K. Alewine and W. Chen, "A review of electrical winding failures in wind turbine generators," *IEEE Electrical Insulation Magazine*, vol. 28, no. 4, 2012.

[15] M. J. Daigle and K. Goebel, "A model-based prognostics approach applied to pneumatic valves," *International journal of prognostics and health management*, vol. 2, no. 2, pp. 84–99, 2011.

[16] Y. Li, S. Billington, C. Zhang, T. Kurfess, S. Danyluk, and S. Liang, "Adaptive prognostics for rolling element bearing condition," *Mechanical systems and signal processing*, vol. 13, no. 1, pp. 103–113, 1999.

[17] Y. Li, S. Billington, C. Zhang, T. Kurfess, S. Danyluk, and S. Liang, "Dynamic prognostic prediction of defect propagation on rolling element bearings," *Tribology transactions*, vol. 42, no. 2, pp. 385–392, 1999.

[18] M. N. Kotzalas and T. A. Harris, "Fatigue failure progression in ball bearings," *Tranctions-American Society of Mechanical Engineers Journal of Tribology*, vol. 123, no. 2, pp. 238–242, 2001.

[19] C. H. Oppenheimer and K. A. Loparo, "Physically based diagnosis and prognosis of cracked rotor shafts," in *AeroSense 2002*, pp. 122–132, International Society for Optics and Photonics, 2002.

[20] L. Bian, N. Gebraeel, and J. P. Kharoufeh, "Degradation modeling for real-time estimation of residual lifetimes in dynamic environments," *IIE Transactions*, vol. 47, no. 5, pp. 471–486, 2015.

[21] P. B. Matej Gasperin and D. Juricic., "Model-based prognostics under non-stationary operating conditions," in *Annual Conference of the Prognostics and Health Management Society*, pp. 1–7, 2011.

[22] H. Liao and Z. Tian, "A framework for predicting the remaining useful life of a single unit under time-varying operating conditions," *IIE Transactions*, vol. 45, no. 9, pp. 964–980, 2013.

[23] F. Zhao, Z. Tian, E. Bechhoefer, and Y. Zeng, "An integrated prognostics method under time-varying operating conditions," *IEEE Transactions on Reliability*, vol. 64, no. 2, pp. 673–686, 2015.

[24] K. Goode, J. Moore, and B. Roylance, "Plant machinery working life prediction method utilizing reliability and condition-monitoring data," *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering*, vol. 214, no. 2, pp. 109–122, 2000.

[25] A. Jardine, P. Anderson, and D. Mann, "Application of the weibull proportional hazards model to aircraft and marine engine failure data," *Quality and reliability engineering international*, vol. 3, no. 2, pp. 77–82, 1987.

[26] A. Jardine, P. Ralston, N. Reid, and J. Stafford, "Proportional hazards analysis of diesel engine failure data," *Quality and Reliability Engineering International*, vol. 5, no. 3, pp. 207–216, 1989.

[27] W. Wang and A. Christer, "Towards a general condition based maintenance model for a stochastic dynamic system," *Journal of the operational research society*, vol. 51, no. 2, pp. 145–155, 2000.

[28] A. H. Elwany and N. Z. Gebraeel, "Sensor-driven prognostic models for equipment replacement and spare parts inventory," *IIE Transactions*, vol. 40, no. 7, pp. 629–639, 2008.

[29] M. Dong and D. He, "A segmental hidden semi-markov model (hsmm)-based diagnostics and prognostics framework and methodology," *Mechanical systems and signal processing*, vol. 21, no. 5, pp. 2248–2266, 2007.

[30] A. Malhi, R. Yan, and R. X. Gao, "Prognosis of defect propagation based on recurrent neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 3, pp. 703–711, 2011.

[31] G. Vachtsevanos and P. Wang, "Fault prognosis using dynamic wavelet neural networks," in *AUTOTESTCON Proceedings, 2001. IEEE Systems Readiness Technology Conference*, pp. 857–870, IEEE, 2001.

[32] S. Zhang, L. Ma, Y. Sun, and J. Mathew, "Asset health reliability estimation based on condition data," 2007.

[33] X. Zhang, R. Xu, C. Kwan, S. Y. Liang, Q. Xie, and L. Haynes, "An integrated approach to bearing fault diagnostics and prognostics," in *American Control Conference, 2005. Proceedings of the 2005*, pp. 2750–2755, IEEE, 2005.

[34] R. B. Chinnam and P. Baruah, "Autonomous diagnostics and prognostics through competitive learning driven hmm-based clustering," in *Neural Networks, 2003. Proceedings of the International Joint Conference on*, vol. 4, pp. 2466–2471, IEEE, 2003.

[35] A. K. Mahamad and T. Hiyama, "Development of artificial neural network based fault diagnosis of induction motor dearing," in *Power and Energy Conference, 2008. PECon 2008. IEEE 2nd International*, pp. 1387–1392, IEEE, 2008.

[36] B. Boashash, *Time-frequency signal analysis and processing: A Comprehensive Reference*. Academic Press, 2015.

[37] M. Cocconcelli, R. Zimroz, R. Rubini, and W. Bartelmus, "Stft based approach for ball bearing fault detection in a varying speed motor," *Condition Monitoring of Machinery in Non-Stationary Operations*, pp. 41–50, 2012.

[38] Z. Feng and M. Liang, "Fault diagnosis of wind turbine planetary gearbox under nonstationary conditions via adaptive optimal kernel time–frequency analysis," *Renewable Energy*, vol. 66, pp. 468–477, 2014.

[39] R. S. Pathak, *The wavelet transform*, vol. 4. Springer Science & Business Media, 2009.

[40] Y. Gritli, L. Zarri, M. Mengoni, C. Rossi, F. Filippetti, and D. Casadei, "Rotor fault diagnosis of wound rotor induction machine for wind energy conversion system under time-varying conditions based on optimized wavelet transform analysis," in *Power Electronics and Applications (EPE), 2013 15th European Conference on*, pp. 1–9, IEEE, 2013.

[41] Y. Guan, M. Liang, and D.-S. Necsulescu, "A velocity synchrosqueezing transform for fault diagnosis of planetary gearboxes under nonstationary conditions," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 231, no. 15, pp. 2868–2884, 2017.

[42] I. Antoniadou, G. Manson, W. Staszewski, T. Barszcz, and K. Worden, "A time–frequency analysis approach for condition monitoring of a wind turbine gearbox under varying load conditions," *Mechanical Systems and Signal Processing*, vol. 64, pp. 188–216, 2015.

[43] W. Wang and W. Zhang, "A model to predict the residual life of aircraft engines based upon oil analysis data," *Naval Research Logistics (NRL)*, vol. 52, no. 3, pp. 276–284, 2005.

Proposed changes in Engineering Students' Final Project & Attachment Management in JKUAT to Enhance University – Jua Kali Collaboration to bridge existing technological

Henry N. Kaosa and Seroni Anyona

***Abstract*—**Apart from being a major employer, Kenya's informal sector which is commonly referred to as *Jua Kali* has produced a lot of useful inventions and improved products that have changed Kenyan lives. Some of these products include airplane prototypes, improved charcoal stoves, wheelbarrows and *money maker* water pumps just to mention a few. However most of these products have failed to attain optimal mass production and improvements due to insufficient technical expertise in carrying out design and optimization, poor research and logistical support, financial constraints and ineffective marketing strategies. For this sector to fully contribute to industrialization of our country, deliberate efforts must be made to support and nurture it. This provides an entry opportunity for institutions of higher learning, like JKUAT, to collaborate and convert the raw indigenous ideas to workable and technological systems for mass production.

***Keywords*—**Industrial collaboration, academia and industry, formal and informal sectors, Jua Kali.

I. Introduction

THIS paper reviews and proposes changes in project management for engineering students in JKUAT. We explore the benefits to be earned if student projects are done outside university premises in collaboration with Jua Kali artisans. Students will be exposed to apprenticeship in the informal sector and get a shot at production using only the equipment available to artisans in the Jua Kali sector. Students may come up with new/ improved designs or fabricate more efficient tools but will work under the supervision of the artisan since the latter is more experienced in the field. An alternative approach is where the artisan does the casting or fabrication as may be required under the guidance of the student who will design, provide intellectual support, pay and write a report on the experience with suggestions for process/ equipment improvement. Ideas and inventions generated at this level can be branded and exposed to the market. This will ensure that the products generated attract funding and raise the much needed income for the sector.

Currently JKUAT has 10 BSc engineering programs under College of Engineering and Technology (COETEC) [3]. These include; Civil Engineering, Electrical and Electronic Engineering, Electronic and Computer Engineering, Telecommunication & Information Engineering, Geomatics Engineering and Geospatial Information Systems (GEGIS), Mechanical Engineering, Mechatronic Engineering, · Mining & Mineral Processing Engineering, Marine Engineering, Agricultural & Bio-systems Engineering. Each year the university produces hundreds of graduates in electronics/ electrical engineering, agricultural, mechanical engineering, mining and mineral processing among others [3].

The college of engineering trains undergraduate and postgraduate students as well as carrying out relevant research. The following are the main objectives:

1. To provide manpower for the development of industry both the public and private sector;
2. To provide relevant manpower for various sectors involved in research in manufacturing, infrastructural development, mining, etc.

The mission of JKUAT is to offer accessible quality training, research and innovation in order to produce leaders in fields of engineering and enterprise development among others to suit the needs of a dynamic world [10]. While the vision of the University is to be a University of Global Excellence in Training, Research and Innovation for Development

Figure 1: Mining engineering students carrying out magnetic survey

[1]H. N. Kaosa, Department of Mining, Materials and Petroleum Engineering, JKUAT (phone: +2540790990466; e-mail: kaosa.henry@jkuat.ac.ke).

[2]A. Seroni, Department of Mining, Materials and Petroleum Engineering, JKUAT (e-mail: sanyona@jkuat.ac.ke).

Figure 2: JKUAT engineering laboratories

Figure 3: JKUAT Electrical engineering laboratories

The training offered in engineering courses lasts for periods of 5 years for undergraduate. In the first three years of the programs, all the students cover common units mainly in mathematics, applied science, fundamental engineering and principles relating to their area of specialization [3]. Each student undergoes mandatory eight-week industry-based practical attachment at the end of the second, third and fourth years of study [3]. The student is usually tasked to write a report of the experience and draw sketches of equipment used [4].

During industrial attachment, the engaging institution/ company is expected to assign a supervisor to the student. This supervisor will instruct and follow up on the activities of the attaché then give recommendation and signature on a daily basis in forms provided by the university. A student is not allowed to graduate until he/she has been to three industrial attachments and spent the required hours.

Target groups for JKUAT's engineering programs include qualified form four graduates, 'A' Level graduates, Diploma holders and holders of other qualifications recognized by the University Senate. Form four graduates are admitted through the Kenya University and College Central Placement Service KUCCPS.

The Engineers Board of Kenya (EBK) is the body mandated to register, regulate and maintains a roll of qualified engineers in Kenya. The EBK was established under Section 3(1) of the Engineers Act 2011. EBK boasts a cumulative registered membership of over 1800 certified engineers. Most of the Engineering courses offered at JKUAT are accredited by the EBK and a good number of JKUAT staff are registered members.

JKUAT was started as a technical college in 1981. It was a cooperative project between the Government of Japan and the Government of Kenya. Since then, this institution has grown tremendously. It is now the leading and most prestigious, most preferred institution of higher learning for engineering enthusiasts in the region, offering training at undergraduate, Masters and Doctor of Philosophy levels [3]. According to [3], the University consists of the College of Engineering and Technology, Faculty of Science, Faculty of Agriculture, the School of Architecture and Technology, School of Human Resource Development and Institute of Computer Science and Information Technology. There are several research institutes that focus on extension activities and scientific inquiry.

Programs in the College of Engineering and Technology fall under three Schools: School of Civil, Environmental and Geospatial Engineering; School of Electrical, Electronic and Information Engineering; School of Mechanical, Manufacturing, Mining, Materials and Petroleum Engineering [3].

I. JUA KALI

Jua Kali is a Swahili word which can be interpreted as 'hot sun'. This refers to Kenya's informal sector where in many cases business is carried out under open air environments exposed to the heat of the sun throughout the day [6]. Jua Kali artisans and traders form the majority of the human capital which powers the economy in urban areas countrywide. According to a research done by J. O. Marwanga [1] there is enough evidence indicating that the majority of the population derives their livelihood from the informal economy. They procure or extract, raw materials, manufacture essential products and distribute various commodities through shops, open-air stalls and hawking. Most Jua Kali activities are concentrated in urban areas though a good number of artisans operate in the rural setting. In Nairobi, Jua Kali artisans are located at Gikomba, Kamukunji, Ziwani, Kawangware, Kariobangi Githurai and Roysambu to mention just a few.

Artisanal apprenticeship may be a result of traditional African craftsmanship which has been around as long as our communities have been handed down from generation to generation. However the artisanal informal sector (Jua Kali) as is known today may have started as early as 1899 at the advent of colonialism. After the completion of the Uganda railway Nairobi and other towns started to mushroom. This also came with the alienation of land and this forced many Africans to migrate to the White settler farms, Nairobi and other urban centres, in search of wage labour. Those who failed to secure wage labour in Nairobi turned to hawking and other marginal activities such as the brewing of traditional liquor and craftsmanship. It is further argued that although a number of Africans acquired artisanal skills from the government departments and mission schools by the inter war period, very

few of them practiced artisanal businesses in Nairobi. Rather, most of them acquired jobs in the formal sector. Therefore, the African pioneer Jua kali artisans were those who acquired their skills informally from the Asian artisans. The Second World War marked the increased exploitation of the Kenyan economy. Most of the Africans who migrated to urban centers, demobilized soldiers after the Second World War and those evicted from white settlement farms after farm mechanization turned to the informal sector to make a living. They engaged in carpentry, metal work, woodwork, bicycle and motor vehicle repair, as well as shoe repair. Evidently, these Jua kali trades, were a direct consequence of the inadequate job opportunities in the waged sector of the economy. At independence in 1963, the influx control measures against African movement into Nairobi were lifted. This in turn accelerated the influx of thousands of the ex-detainees and other Africans who had been repatriated from Nairobi during the Emergency, educated and semi-educated job seekers. It is against this background that many Kenyans were pushed into the Jua kali sector after independence into the 1980s and 1990s.

The Jua Kali industry constitutes shop owners, engineers, craftsmen, apprentices and road-side traders etc. Their activities include repair or production oriented activities in motor vehicle repair, carpentry and woodwork. Areas of specialization have broadened since M. Walsh's [2] publication and has incorporated very many additional areas ranging from electronics to beauty and cosmetics. The products include amplifiers, inverters, wheelbarrows, furniture, kitchenware, metal suitcases for students (*sanduku*) and clothes, which are sold locally and regionally in the various market centers. Main means of distribution is through the traditional commercial supply chain and infrastructure.

Figure 4: Wheelbarrows & Metallic Boxes made by Jua Kali

Figure 5: Improved Charcoal stove by Jua Kali artisans

Figure 6: An assortment of Jua Kali products

The potential of the informal sector and the Jua Kali sector in particular as a vehicle for job creation and a tool for wealth generation was first recognized by ILO in 1971 [5]. More attention was focused on the movement in 1986 triggering the creation of the Ministry of Technical Training and Technology by the Government of Kenya. In more recent years, the definition of Jua Kali has evolved to mean small scale operators who are more limited by resources than by choice.

About 89.72 per cent of the 832,900 new jobs created in 2016 were from the informal sector. The 2017 Economic Survey by the Kenya National Bureau of Statistics showed that the informal sector, commonly referred to as Jua Kali, generated 747,300 jobs while the remainder 85,600 were created by the formal sector.

II. Literature Review

A. Kenyan Academia's Current Project Management and Apprenticeship/ Industrial Attachment

A lot of studies have been carried out regarding the Jua Kali sector by many academicians and researchers from various Kenyan universities. These include JKUAT, UoN, KU, Moi, etc. Most of these studies have focused on various thematic areas of the sector including government policy, marketing, value addition, technological improvements, organization, training, intellectual property etc. The authors referenced in this paper have demonstrated understanding of the sector's challenges. Kathuru Joseph Gitari (UoN) studied The Emergence and Development of the Informal Sector 1908-2009 in Eldoret. J. O. Marwanga,did a study on "The impact of informal economy on employment creation," the Case of

Kamukunji Jua Kali artisans in Nairobi, Kenya [1], Berlin, J. Sullivan ed., cried out a study titled "Under the sun or in the shade? Jua Kali in African countries and S. M. Maina, R. W. Rukwaro, [2] M.Walsh,on "Education, training and the informal sector in Kenya" and W. H. Onyango, wrote a paper on "Infusing design in the Jua Kali production processes" among others.

The evidence investigated by M. Walsh [2] shows that the informal sector can absorb many more apprentices than are trained at present. According to M. Walsh the total number of trainees could be more than doubled simply by filling up the existing underutilized capacity. It is noted that the pattern of underutilization seems to have arisen because entrepreneurs are experiencing difficulty in recruiting capable apprentices, a reflection of the fact that many of them are no longer happy to take on trainees primarily for social reasons, but are beginning to put more economic considerations first. M. Walsh proposes new uses of existing training institutions acknowledging that the existing public training institutions possess the capacity, at least in theory, to service the informal sector in skill upgrading, product diversification, quality improvement, and research for appropriate technology.

B. Short-comings of Formal Education System

In Kenya, according to UNEVOC Berlin, J. Sullivan ed [5], the majority of the unemployed, characterized by people without work who are seen idling around or actively looking for work, are the young and relatively more educated. It is estimated that the informal sector will create about 100,000 jobs per year if critical training programs which are market-driven are created [5]. It is also noted that despite high demand for education and major investments in this sector by Sub-Saharan African countries, the region's education systems are not producing the intellectual and human resources required for sustainable economic growth and political stability. Its recommends that "The supply of technicians as well as of craftsmen and artisans should be augmented through 'on-the-job' and 'in-service training', a move that should involve both government department and the private sector.

III. JUSTIFICATION

The human resource produced from Kenyan universities has not been fully and effectively utilized in powering industrialization geared towards substantial economic development for the country. According to J.O. Marwanga [1] and UNEVOC Berlin, J. Sullivan [5], there is urgent need to come up with proactive radical solutions to this problem so as to create employment opportunities that match the high rate of growth of the labor force in Kenya. Long-term collaboration between JKUAT and the Jua Kali sector with regards to innovation has a number of benefits:

1. It will help link academia and researchers to the informal and indigenous sector creating the much needed synergy that will utilize the strengths of two sectors to propel industrialization.
2. The Jua Kali sector may be compared to the pre-industrial Europe cottage industry which formed the foundation for industrialization. This sector is at its infancy and therefore provides opportunities for innovations, improvisations and research.
3. Unlike the formal sector the Jua Kali sector is not riddled with rigid bureaucracy and red tape which inhibit creativity and spontaneous thinking and therefore the Jua Kali sector may accommodate more mistakes providing the opportunity to test new ideas. This may also help improve the quality of the sector's products and change the negative public perception about the sector.
4. Most artisans engaged in the Jua Kali industry do not have sufficient education required to document and protect their intellectual property. This has led to lose of indigenous Kenyan knowledge, innovations and ideas to other countries. So, University-Jua Kali collaboration will help seal this loophole.
5. Currently there is a critical shortage of vacancies for student attachment in the formal sector. So opening up and accepting the informal sector as an alternative provider for attachment with create limitless opportunities.

A. Optimal Utilization of the Academic Knowledge and Creativity of Students

In Kenya, engineering knowledge and expertise is diverse. There exists traditional craftsmanship passed down from parent to offspring, *Jua Kali* craftsmanship in urban areas and academically qualified engineers.

When engineering is supported with quality management, theoretical research, design, standardization, financing and government policy support, success becomes inevitable. This can be witnessed from the developments in the West and Asian countries throughout the last century and currently.

Students are usually gifted with innovative talent and JKUAT offers the convenience of having experts from a wide range of interdependent academic disciplines within the same institution. This provides an ideal place for catalyzing the creation of complete industrial systems by incubating ideas from various departments. Experts from the informal sector can provide the much needed experience in tool handling, efficient methods, no-go's and much more.

B. Industrial attachment and students' final project

Apart from the theoretical classes, mentioned elsewhere, engineering students at JKUAT are required to undergo practical attachment within and without the university. In addition they also undertake an individual final year project. A student will undergo a mandatory practical training attachment, each of eight weeks, on at the University at the end of the second year and the others at the end of third and fourth years of study including units of practical training in the engineering workshops during the academic semesters. A student must undertake, in their final year of study, a suitable engineering project under supervision of a member of academic staff. All this is geared towards ensuring that the student acquires the basic engineering skills and innovative attitude.

To appreciate the dynamics of attachment in JKUAT, two departments have been sampled namely Mechanical Engineering Department and Mining, Materials & Petroleum Engineering (MMPE) Department for the year 2017-2018. The table below shows the frequently visited companies for attachments by students and the activities that they are involved in.

Attachment records for 4th year MMPE Students 2017-18

Name of company	No of students	Activities
Ministry of Mining; Mines and Geology Department	2	• Map analysis and digitization. • Geophysical exploration methods • Explosives and drilling • Assaying • Geochemistry lab and equipment (atomic absorption and spectrometer, gold analysis). • Lapidary section(jaw crusher and pulveriser) • X-ray fluorescence
East Africa Portland Cement	3	**KURKUR QUARRY** • Quarry operations, Crushing operations and collection of cycle time data • Tracking and mapping quarry boundaries using GPS mapping tools • Limestone crusher inspection • Laboratory analysis samples **BISSEL QUARRY** • Safety and Health policy • Drilling operations • Blasting of drill holes **KABINI QUARRY** • Location of drill holes using GPS • Maintenance of diesel crusher • Clinker manufacturing and kiln operation • Cement milling and packaging
City stone Quarry	9	• Safety and Maintenance operations. • Drilling. • Stone cutting. • welding • Drilling holes for blasting, charging, stemming and blasting. • Operation of a wheel loader.
National Water Corporation	1	• Beams and columns. • Well design and construction. • Hydrological surveying and stratigraphy. • Drill pipes, bits and rotating equipment. • Hoisting equipment. • Rig instrumentation system • Noise control.
Wazaire Gems Company	1	**Study of gemstones** • Cut and rough gemstones, • Naming and measuring gems (Trade names and misnomers) • physical and chemical properties of gems • Study of magnification using loupe and gem microscope. • Grading of gems

Mechanical Engineering Department attachment record 2017-18

Company/Institution	No.of Students	Activities
GIZ Westlands	2	• Repair and Maintenance • Installations • Design
DT Dobie, Lusaka rd	1	• Repair and Maintenance • Installations • Design
East African Breweries Limited, Ruaraka	2	• Repair and Maintenance • Installations • Design
British American Tobacco LIMITED. Likoni RD	1	• Repair and Maintenance • Installations • Design
Numerical Machining complex	1	• Repair and Maintenance • Installations • Design
Nairobi Bottlers Limited, North airport rd	1	• Repair and Maintenance • Installations • Design
Unga farmcare EA Ltd, Dakar RD. Ind. Area	1	• Repair and Maintenance • Installations • Design
East African Portland cement, Athi river	1	• Repair and Maintenance • Installations • Design
Air Kenya Express LTD, Wilson Airport	1	• Repair and Maintenance • Installations • Design
JB Motors, Industrial area	1	• Repair and Maintenance • Installations • Design
CMC Nairobi, Lusaka rd/ Heavy commercial vehicles	2	• Repair and Maintenance • Installations • Design
Kahawa Barracks	2	• Repair and Maintenance • Installations • Design

Kapa Oil Refineries Ltd, Mlolongo	1	• Repair and Maintenance • Installations • Design
Subaru Kenya, Industrial area	1	• Repair and Maintenance • Installations • Design
Central Glass Industries, Kasarani	1	• Repair and Maintenance • Installations • Design
Kenya Airports Authority, Kisumu Airport	1	• Repair and Maintenance • Installations • Design
JEEP Motors TATA Kericho	1	• Repair and Maintenance • Installations • Design
Nzoia sugar company	1	• Repair and Maintenance • Installations • Design
Mumias Sugar Co.	1	• Repair and Maintenance • Installations • Design

From the sampled information it is observed that: (1) the preferred companies are all formal establishments. There is no informal establishment in the list. (2) The students are involved in regular routine operations. (3) Operations involving systems design and innovation are not open to students on attachment. (4) The existing systems & structures are rigid and not able to accommodate any new untested innovative ideas from students.

While understanding the existing operational systems of a company is important, development of key engineering skills at the earliest opportunity is critical. This requires companies to be prepared to take risks and create room and space for innovations within their existing structures. This will involve allowing students to participate in designs of new models, piloting and implementation. It is in this regard that it's proposed in this paper that the informal Jua Kali Sector be considered as an alternative for attachments and internships.

Apart from providing more chances for attachment, the informal sector will have a number of added advantages. It's less restrictive and more flexible in accommodating new ideas. It will be possible to take risks without adverse impacts on related systems. It will be easier for the student to own intellectual property of any innovation. The outcome of any innovation will be felt immediately given the fact that in the Jua Kali industry the innovator is in direct contact with consumers. Innovations developed under these circumstances can be easily and cheaply replicated and hence will motivate entrepreneur desire in the student. There are greater growth possibilities in the Jua kali sector going forward than the formal sector.

C. STUDY HYPOTHESIS

This study asserts that the informal Jua Kali sector will provide the most appropriate environment for identifying, developing, nurturing, and employing young innovative skilled engineers from JKUAT using the student attachment and project format than the formal sector. This will position students for the future in this most promising sector.

D. OBJECTIVE

The key objective of this study is to explore and illustrate the merits of University-informal sector collaboration and the opportunities the informal sector can provide for engineering innovations.

- To encourage student attachments in the informal sector and increase the vacancy available for attachment
- To promote the utilization of the growth potential of the informal Jua Kali sector for innovation and experimentation
- Integrate IT in the informal sector
- Help improve the quality of products and services provided by the Jua Kali informal sector and make them competitive in the local and international market.

E. METHODOLOGY

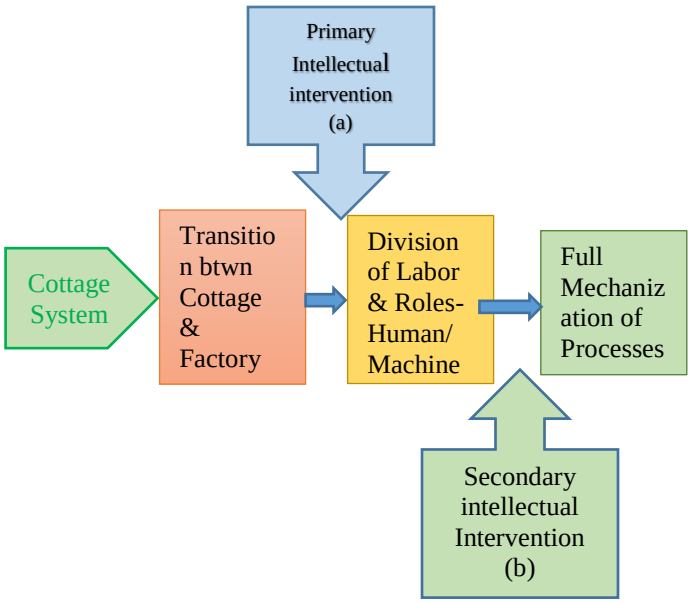
This study was carried out by doing desk review of existing secondary data and related researches and documentation. However a substantial amount of information was acquired by unstructured interviews of the Jua kali artisans and university staff as well as through observation.

IV. INFUSION OF INTELLECTUAL KNOWLEDGE IN THE JUA KALI SECTOR: THE WAY TO THE FUTURE

Past experience from other regions such as Europe show that there are four systems leading to industrialization namely; Family System, Handicraft System (Guild System), Domestic System or Cottage System and Factory System. The Kenyan Jua Kali sector currently represents the cottage system which is at the threshold of the indigenous factory system.

Four systems leading to industrialization

Development of the Cottage industry to Manufacture



From the flow chart above and according to this paper cottage or informal sector industry develops in four stages to full industrialization. The initial cottage activities undergo a transition period in which there is both human and machine labor and some parts of the processing are done under cottage while the other parts are done by factory processes. This will then mature to a stage of division of labor in which specific processes that produce specific products are concentrated in one location to supply the rest of the production processes. The final stage is that of full mechanization where over three quarters of the production is done by machine.

Based on this Jua Kali development conceptual model there are key entry points for academic intervention. These may be referred as primary intervention (a) and secondary intervention (b). At the primary stage the key interventions will include design of factory processes, design of packaging, acquisition of intellectual property, quality assurance, marketing and health and safety among others. The secondary stage of intervention may well be similar to the first stage however at this point there will be need for advanced mechanization including digital systems, robotics and smart technology.

V. CONCLUSION

As the Jua Kali sector continues to develop there will be limitless opportunities for academic and research communities to be part of this dynamic sector. This development will provide room and space for experimentation, innovation and creativity. The earlier this opportunity is sized the better. The proposed student attachment model can and will be a stepping stone in the right direction.

Jua Kali artisans have also accumulated undocumented skills throughout their many years of innovations and improvement, which is guided mainly by market demand enforced by the need to earn a living in the modern financial system. Innovations need to be studied, catalogued and documented to facilitate future developments.

Being an accomplished academician should not separates an individual from the larger society because of the way in which the science which drives engineering discipline strives to cope with first world trends. A desirable alternative would be studying and working to improve local industry while learning from the mistakes of developed countries.

The mistakes made by developed nations include manufacturing products with unchecked levels of carbon emission – which is the main contributor to the current problem of global warming, lead poisoning, overproduction with a focus on maximization of profits rather than continuous development targeted at fulfilling existing needs

Our education systems should have the power to return an improved individual to the society, one who not only understands the society's inner workings and technologies but is also capable of adding value by building upon the existing systems. For this feat to be achievable, local traditional and industrial products must be cataloged, their origins and developments taught to students who are expected to make improvements and other innovations with an aim of meeting existing needs with greater optimized efficiency.

Even as we encourage the education of our children, it becomes essential that we find a way of utilizing the knowledge and skills available in the communities outside the education system

It is the submission of this paper that efforts be made to adopt the current attachment and final student project requirements to suit the informal Jua Kali sector and encourage students to venture into the sector.

REFERENCES

[1] J. O. Marwanga, "The impact of informal economy on employment creation," the Case of *Kamukunji Jua Kali* artisans in Nairobi, Kenya, International journal of economics, commerce and management, vol. III. issue 6, United Kingdom, June 2015, pp. 854, 860–862, 873.
[2] M. Walsh, "Education, training and the informal sector in Kenya", October 1991, pp. 18–19, 23.
[3] JKUAT COETEC brochure
[4] JKUAT Faculty of engineering students external attachment Log-Book
[5] UNEVOC Berlin, J. Sullivan ed., "Under the sun or in the shade? Jua Kali in African countries", National policy definition in technical and vocational education: beyond the formal sector, international project on technical and vocational education, Nairobi, 1998, pp. 17, 39–41.
[6] S. M. Maina, R. W. Rukwaro and W. H. Onyango, "Infusing design in the Jua Kali production processes", IOSR journal of humanities and social science, vol 23, issue 3, ver II, March 2012, pp 1,
[7] JKUAt Intake Advertisement: http://host.nationmedia.com/JKUAT-May-September-2016-Intake.pdf, accessed 6th April 2018, 9:53 am
[8] Corporate Communications Office, "New students report to JKUAT campuses", http://www.jkuat.ac.ke/new-students-report-to-jkuat-campuses/, accessed on April 6th 2018 at 10:23 am
[9] Engineers Board of Kenya, "Registered professional engineers", http://ebk.or.ke/registered-professional-engineers/ accessed on 6th April 2018, 10:45 am
[10] JKUAT, June 2017 Graduation Booklet
[11]

Continuous decontamination of metal-polluted mine water using engineered hybrid adsorbent

AM. Muliwa, MS. Onyango, A. Maity and A. Ochieng

***Abstract*-** Mining industries contribute enormously to water pollution through discharge of effluents contaminated with metals. Metals-polluted water is a threat to aquatic and human lives as well as the general ecosystem. Numerous conventional treatment methods are available for the removal of metals from mine water, but majority of them are costly, inefficient for trace metal concentration, and generate voluminous secondary sludge. Therefore, there is need for alternative low-cost novel technologies capable of reducing metal concentration in water to acceptable levels. Adsorption technology is increasingly receiving preference because it is simple in design, requires low initial cost, easy to operate, can remove contaminants in trace levels and the possibility to develop and employ wide variety of adsorbents. This study, therefore, explored the removal of manganese [Mn (II)] from aqueous solution using a bentonite/metal oxide hybrid ($B/MeO_x/H$) hydrogels adsorbent packed in a fixed-bed reactor. The adsorption performance was examined by breakthrough behaviors under various experimental conditions such as bed height and influent flow rate. It was found that increase in bed height resulted to increase in both breakthrough and saturation times, while it was opposite with influent flow rate. The breakthrough curves were characterized by slightly flatter breakthrough curves for large bed depth and low flow rate. The overall adsorption performance of the fixed-bed column highly depended on the operating parameters. The experimental breakthrough data was sufficiently described by Yoon-Nelson model. Bed depth service time (BDST) predicted well the breakthrough times for different flow rates. The findings demonstrate that $B/MeO_x/H$ hydrogel adsorbent could effectively polish mine water laded with trace concentration of Mn (II).

***Keywords*—**Adsorption, Breakthrough, Hybrid, Mine water

1. Introduction

Water pollution by manganese (Mn) mainly results from the disposal of partially treated effluents from electroplating, mining, coal and oil burning battery manufacturing, chemical manufacturing, and metal alloying industries, among others, into municipal sewers [1]. Elevated concentration of Mn is detrimental to both living organisms and the general environment. Under normal environmental conditions, the divalent form [Mn (II)] is the most stable and predominates over a wide range of pH. Due to Mn (II) toxicity and non-biodegradability, there is urgent need for innovative technologies capable of removing Mn (II) from industrial effluent to acceptable standards [2]. Conventional treatment methods of Mn(II)-contaminated waters include chemical precipitation, ion-exchange, membrane filtrations, biological methods and adsorption etc. Precipitation is highly commercialised but, the requisite to raise the pH above 10, so as to comply with the discharge limits, is a problem for discharging [3].

Adsorption technique has attracted a lot of attention because it is simple, non-toxic, and suitable for trace level contaminants concentrations and does not generate secondary waste. Moreover, adsorption technology is flexible and can be integrated into an existing treatment system to achieve complete metal ions removal [4]. Several adsorbents have been reported for metal ions removal from aqueous solutions. However, there is growing interest in low-cost adsorbent (LCAs) materials as they are cheap, surface-modifiable to enhance adsorption capacity and selectivity, and they can be regenerated for re-use [5]. Among these LCAs, bentonite has shown promising results with regard to environmental clean-up, thanks to its high cation exchange capacity, layered structure, increased surface area, and swelling characteristic which enables them retain contaminants within the structure [6]. However, bentonite still exhibit low capacities for metal ions, stickiness and poor chemical and mechanical properties. In order to improve the adsorption performance of bentonite, it is important to physically or chemically modify their surfaces [7].

Metal oxides are widely known as efficient scavengers of metal ions [8] and could serve as modifying agents of bentonite. Metal oxides would not only provide reactive surface for metal removal, but also densify bentonite and improve the dispersion characteristics of the resultant adsorbent. However, both bentonite and metal oxides are usually available in powder form, making them unsuitable for application in continuous fixed-bed adsorption. Fixed-bed columns packed with powdered adsorbents suffer from high pressure drops, thus particles of appropriate sizes should be used. Consequently, in this study, manganese oxides (MeO_x) were immobilized on bentonite (B) surface and bound on chitosan to produce hydorgels for Mn (II) adsorptions. The

A.M. Muliwa, MS Onyango, Department of Chemical, Metallurgical and Materials Engineering, Tshwane University of Technology (corresponding author phone: +27818445644; fax: +27123823550; e-mail: amuliwa@yahoo.com).

A. Maity, DST/CSIR National Centre for Nanostructured Materials, Council for Scientific and Industrial Research, Pretoria.

A. Ochieng, Centre for Renewable Energy and Water, Vaal University of Technology, South Africa

choice of MeO_x was guided by the fact that they have demonstrated superior adsorption performance towards metal ions than other metal oxides, attributable large porosity, high catalytic activity and large surface areas [9]. Furthermore, chitosan being biocompatible and non-toxic polymer was chosen as a binder. The resultant adsorbent (B/MeOx/H) was examined for the sequestration of Mn (II) using a continuous fixed-bed operation. The influence of various process variables on Mn (II) breakthrough curves was explored. Finally, experimental breakthrough data was fitted with Yoon-Nelson, Adams-Bohart and bed depth service time (BDST) models to predict the performance under different conditions.

2. Materials and Methods

2.1. Adsorbent preparation and characterization

The hydrogels were prepared as follows: First, 0.2M permanganate solution was prepared and heated to boiling at 90ºC. Then, 2 g of bentonite powder was added and the mixture was stirred vigorously for 10 min. Next, 5 mL hydrochloric acid (32% m/v) was titrated into the mixture and a brown suspension was formed. The mixture was further maintained at 90ºC for 1 h, after which it was allowed to cool naturally. Thereafter, the brown solids were vacuum filtered and dried at 100ºC for 24 h. Separately, 1g of chitosan flakes was dissolved in 100 mL of 5% glacial acetic acid until a gel was formed. To this, 1 g of the brown-dried powder from the first experiment was added and the mixture was stirred for 30 min to achieve homogeneity. The resulting mixture was then added dropwise into 0.2M NaOH solution using a syringe tube. Consequently, spherical hydrogels were formed through coagulation and were left in the NaOH solution overnight under soaking. Finally, the brown hydrogels were recovered and washed thoroughly with plenty of water, air-dried and labelled B/MeO_x/H, awaiting further use.

2.2. Continuous decontamination studies

Figure 1 shows the fixed-bed experimental set-up used in this study.

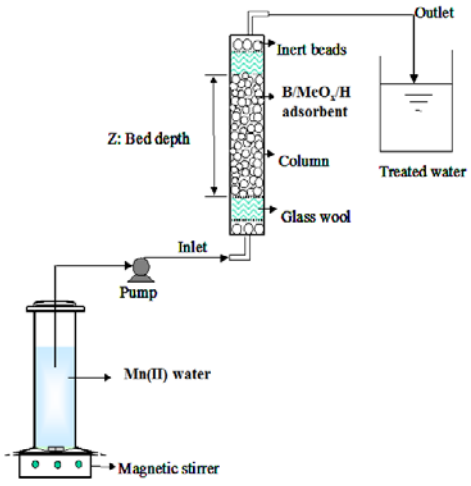


Fig. 1. Fixed-bed adsorption column experimental set-up

A Perspex column with 2 cm internal diameter and 30 cm length was used for experiments. B/MeOx/H hydrogels were placed in the middle of the column tube supported on either side by inert glass beads and glass wool. Mn (II) water was pumped in an upward fashion using a peristaltic pump and effluent samples were collected at predetermined time intervals.

The effect of process parameters such as bed depth (20, 30 and 40 mm) and influent flowrate (1.67, 3.33 and 5.1 mL/min) on breakthrough curves (BTC) was investigated. For the effect of bed depth, the influent concentration of Mn (II) and flow rate were kept constant at 50 mg/L and 3.33 mL/min, respectively. On the other hand, the effect of influent flow rate was explored by fixing bed depth at 20 mm and influent concentration at 50 mg/L. In both experimental runs, samples were collected periodically from the column exit and were analyzed for Mn (II) using inductively coupled plasma optical emission spectrometer (ICP, 9000 Shimadzu). Then breakthrough curves were constructed by plotting C_t/C_o versus time, *t*, whereby C_t is the effluent concentration at any given time and C_o is the influent concentration.

3. Results and Discussion

3.1. Adsorbent characterization

Figure 2 (a) shows SEM image of B/MeOx/H hydrogel bead at x50 magnification while Fig. 2 (b) shows the corresponding EDX spectra. The bead appears almost spherical with a rough surface, which an important feature in adsorption application as it exposes large contact surface area between the adsorbate and the adsorbent. The EDX spectra confirms the presence of all elements representing the individual components making up the hydrogel. Si, K, Al, Ca and Na, represent bentonite, Mn and O peaks confirms the presence of manganese oxides.

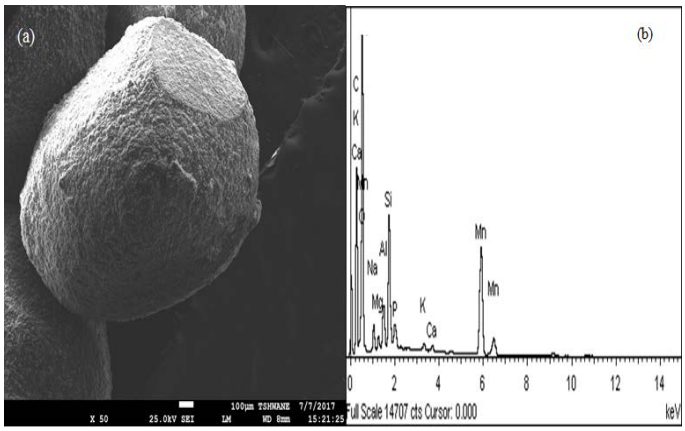


Fig. 2. Scanning electron microscope images (a) and energy dispersive X-ray spectra (EDX) of B/MeO_x/H adsorbent

3.2. The effect of process variables on Mn (II) breakthrough curves

Process variable are known to affect the performance of any adsorption system. Therefore, in this study, the effect of bed depth and influent flow rate on Mn (II) breakthrough curves

were investigated. Figure 3 displays breakthrough profiles obtained for different bed depths. As can be seen, early breakthrough times are evident for shorter bed depths while delayed breakthrough times are characterized by longer bed depths. The same scenario is exhibited for bed saturation times. These results were ascribed to increased active sites available for Mn (II) adsorption for longer bed depths and vice versa [10]. Moreover, for a fixed flow rate, increase in bed depths results in longer residence time which enhances contact time between the adsorbate and the packed hydrogel material. Considering the nature of breakthrough behavior, 20 mm bed depth was utilized for further adsorption experiments.

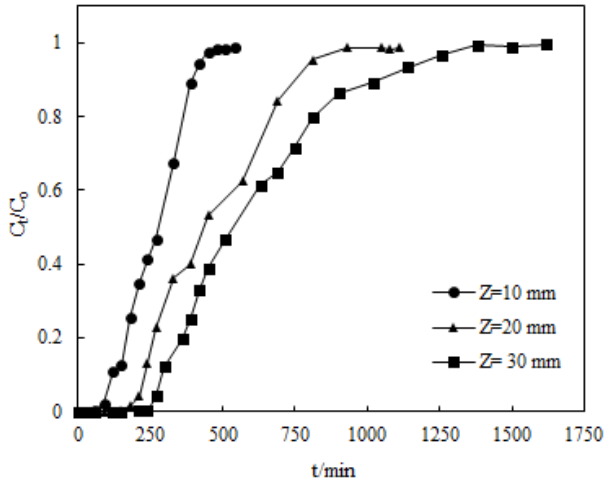


Fig. 3. Effect of bed depth on Mn (II) breakthrough behavior (C_o= 50 mg/L, Q=3.33 mL/min

The effect of influent flow rate on breakthrough curves was also explored and results are presented in Fig. 4.

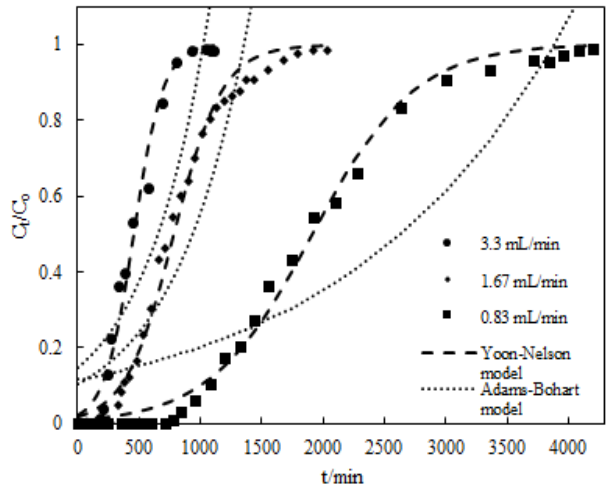


Fig. 4. Effect of influent flow rate on breakthrough curves and Yoon-Nelson and Adam-Bohart model fits (Z= 20 mm, C_o= 50 mg/L)

Contrary to the results on bed depth, increasing influent flow rate resulted in short breakthrough time while a decrease in flow rate produced delayed breakthrough curves. Further, for small flow rate, the breakthrough curves appeared elongated while at higher flow rates, sharper profiles were realized [11]. Early breakthroughs were attributed to less contact time and vice versa. These results indicate that less flow rates favor Mn (II) removal by B/MeOx/H hydrogels.

The experimental breakthrough data was further fitted with Yoon-Nelson model (Eq. (1)) and Adams-Bohart model (Eq. (2)) [12] to predict the adsorption capacity and possible mechanisms.

$$\ln\left(\frac{C_t}{C_o - C_t}\right) = k_{YN}t - k_{YN}\tau \quad (1)$$

where k_{YN} (min^{-1}) is the rate constant, τ (min) is the time required for 50% breakthrough, t (min) is the operation time.

$$\ln\left(\frac{C_t}{C_o}\right) = k_{BA}C_o t - k_{BA}N_o\frac{Z}{U} \quad (2)$$

where k_{BA} (L/mg.min) is the kinetic constant, N_o (mg/L) is the saturation concentration, Z (cm) is the bed depth, and U (cm/min) is linear velocity inside the fixed-bed.

For the two models, plots of the linear equations (not shown) were used to extract parameters from the slope and intercept, after which the obtained parameters were used to predict the non-linear fittings. Figure 4 shows sample model fitting with the influent flow rate. It is clear that the Yoon-Nelson model followed the breakthrough data closely while Adams-Bohart model displayed great deviations. The regression coefficients of Yoon-Nelson model were higher (R^2>0.98) than those of Adams-Bohart model. These results confirmed that Yoon-Nelson model described the breakthrough data satisfactorily. Consequently, it was concluded that the rate of decrease in the probability of an adsorbate molecule is proportional to the probability of adsorbate adsorption and the probability of adsorbent breakthrough on adsorbent. These results are consistent with others reported in literature [13].

3.3. Scale-up study

Scale-up test were conducted to determine if the bench-scale data obtained could accurately predict the performance of a scaled-up fixed-bed unit at different operational parameters. To achieve this, the bed depth service time (BDST) model was applied for predictions because it is the most widely acceptable model for this purpose [14]. The linear equation of BDST model is given as:

$$t = \frac{N_o Z}{C_o U} - \frac{1}{k_B C_o}\ln\left(\frac{C_o}{C_t} - 1\right) \quad (3)$$

where N_o (mg/mL) is the capacity, k_B (mL/mg.min) is the rate constant, U is the linear velocity (cm/min), and Z (cm) is the bed depth. A plot of t versus Z should yield a straight line, from which values of N_o and k_B can be extracted from the slope and intercept.

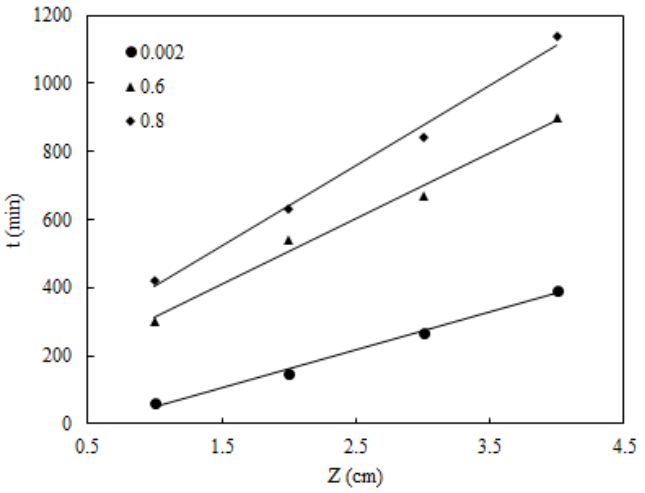


Fig. 5. The iso-removal lines at different C_t/C_o ratios

Figure 5 shows the iso-removal lines for different C_t/C_o ratios (0.002, 0.6, and 0.8), obtained after plotting Eq. (3) while Table 1 gives the summary of parameters calculated from the BDST plots. All iso-removal lines were linear with high regression coefficient values (R^2>0.98) which suggested better fitting of the experimental data with the BDST model. The k_B values decreased with increased effluent concentration while N_o values increased with effluent Mn (II) concentration. Moreover, the slope and intercept were significantly affected by change in effluent concentration. Thereafter, BDST parameters were used to predict saturation time at breakthrough for 75 mg/L influent concentration at constant flowrate (3.33 mL/min) and bed depth (20 mm). The predicted saturation time was 108 min, which was very close to the experimentally obtained saturation time (95 min). The results confirm the suitability of the BDST model in predicting the performance of a fixed-bed column packed with B/MeOx/H adsorbent material for Mn (II) clean-up.

Table 1. A summary of calculated parameters from the BDST model (C_o= 50 mg/L, Q= 3.33 mL/min)

C_t/Co	a (min/cm)	b (min)	k_B x10^{-3} (L/mg.min)	N_o x10^3 (mg/L)	R^2
0.002	111	60	2.07	2.941	0.9956
0.6	193	-120	0.07	5.114	0.9882
0.8	237	-165	0.17	6.281	0.9914

4. Conclusions

The present study investigated the ability of B/MeO$_x$/H hydrogel beads to continuously decontaminate Mn (II)-polluted water in a fixed-bed column. SEM and EDX characterization confirmed rough morphologies and the presence of all individual components. Increasing bed depth increased the breakthrough time and bed saturation time while it was the opposite with the flow rate. Linear regressions indicated that the Mn (II) breakthrough data can sufficiently be described by the Yoon-Nelson model. Furthermore, the predicted performance according to bed depth service time (BDST) model were in agreement with experimental values. This confirmed that BDST model can be used for scale-up studies under different operational conditions. The results provides important insights into the remediation of Mn (II)-contaminated water using B/MeO$_x$/H hydrogel adsorbent.

Acknowledgement

Authors acknowledge Rand Water South Africa through the Rand Water Chair in Water Utilization for financial support towards this project.

References

[1] A. Omri, M. Benzina, "Removal of manganese(II) ions from aqueous solutions by adsorption on activated carbon derived a new precursor: Ziziphus spina-christi seeds," *Alexandria Eng. J.*, vol. 51, pp. 343-350, 2012.

[2] A.M. Silva, E.C. Cunha, F.D. Silva, V.A. Leao, "Treatment of high-manganese mine water with limestone and sodium carbonate," *J Cleaner Prod.*, vol. 29-30, pp. 11-19, 2012.

[3] N. Li, J. Zhao, Z. Jiang, B. Cao, Y. Lu, D. Shan, Y. Zhang, "Continuous manganese(II) ions removal from aqueous solutions using rice husk ash-packed column reactor," *Desalination Water Treat.*, vol. 21, pp. 916-924, 2016.

[4] F. Ji, C. Li, J. Xu, P. Liu, "Dynamic adsorption of Cu(II) from aqueous solution by zeolite/cellulose acetate blend fiber in fixed-bed," *Colloids Surf. A: Physicochem. Eng. Aspects*, vol. 434 (2013), pp. 88-94, 2013.

[5] A. Tripathi, M.R. Ranjan, "Heavy Metal Removal from Wastewater Using Low Cost Adsorbents," *Bioremed. Biodegrad.*, vol. 6 (2015). DOI: 10.4172/2155-6199.1000315

[6] S.A. Al-Jlil, "Removal of heavy metals from industrial wastewater by adsorption using local bentonite clay and roasted date pits in Saudi Arabia," *Trends Appl. Sci. Res.*, vol. 5, pp.138-145, 2010.

[7] K.G. Akpomie, F.A. Dawodua, "Potential of a low-cost bentonite for heavy metal abstraction from binary component system," *Beni-Suef Univ. J. Basic Appl. Scie.*, vol.4, pp. 1-13, 2015.

[8] A.M. Mahmoud, F.A. Ibrahim, S.A. Shaban, N.A. Youssef, "Adsorption of heavy metal ion from aqueous solution by nickel oxide nano catalyst prepared by different methods," *Egypt. J. Petrol.*, vol. 24, pp. 27-35, 2015.

[9] D. Zhou, D.G. Kim, S.O. Ko, "Heavy metal adsorption with biogenic manganese oxides generated by Pseudomonas putida strain MnB1," *J. Ind. Eng. Chem.*, vol. 24, pp. 132-139, 2015.

[10] J.A. Alexander, A. Surajudeen, E.N.U. Aliyu, A.U. Omeiza, M.A.A. Zaini, "Multi metals column adsorption of lead(II), cadmium(II) and manganese(II) onto natural bentonite clay," *Water Sci. Technol.*, vol. 76, pp. 7-8, 2017.

[11] F.W. Sousa, A.G. Oliveira, J.P. Ribeiro, M.F. Rosa, D. Keukeleire, R.F. Nascimento,"Green coconut shells applied as adsorbent for removal of toxic metal ions using fixed-bed column technology," *J. Environ. Manage.*, vol. 91, pp. 1634-1640, 2010.

[12] S. Vilvanathan, S. Shanthakumar, "Continuous biosorption of nickel from aqueous solution using Chrysanthemum indicum derived biochar in a fixed-bed column," *Water Sci. Technol.*, vol 76, pp. 1895-1906, 2017.

[13] M. Bhaumik, K. Setshedi, A. Maity, M.S. Onyango, "Chromium(VI) removal from water using fixedbed column of polypyrrole/Fe_3O_4 nanocomposite," *Sep. Purif. Technol.*, vol. 110, pp. 11–19, 2013.

[14] A. Saravanan, P.S. Kumar, M. Yaswanthraj, "Modeling and analysis of a packed-bed column for the effective removal of zinc from aqueous solution using dual surface-modified biomass," *Part. Sci. Technol.*, 2017. https://doi.org/10.1080/02726351.2017.1329243.

Power System Congestion Management by Generator Active Power Rescheduling using Cuckoo Search Algorithm.

Irungu G. Wangunyu[a], David K. Murage[b] and Peter K. Kihato[c]
[a,b,c]Department of Electrical and Electronic Engineering, Jomo Kenyatta University of Agriculture and Technology, Kenya

Abstract— Power system restructuring has resulted in an increase in complexity of the power flow problem. There has been an increase in the number and volume of transactions from the various market participants as they try to make full utilization of the existing resources for profit maximization. In addition, there is a slow rate in construction of new transmission lines due to environmental, right-of-way and economic hurdles. As a result, transmission systems are at most times operated close to their thermal limits leading to frequent occurrence of network congestion. Congestion imposes a barrier to trade in electrical power and poses a threat to the secure, reliable and economic operation of a power system. Hence, congestion management is a fundamental transmission network management problem that the Independent System Operator (ISO) has to frequently address in an open electricity market.
Generator active power rescheduling is the most popular transmission network overload alleviation technique since it offers ease of control at no additional capital cost. In solving the congestion management problem by generator rescheduling, the aim is to alleviate line overload with minimum rescheduling cost while satisfying the power system equality and inequality constraints. The proposed congestion management problem is formulated as a non-linear, non-convex and highly constrained optimization problem. Thus, solution using Swarm Intelligence (SI) algorithms is suitable. This work studies the effectiveness of Cuckoo Search Algorithm (CSA) in solving the congestion management problem in a pool-based electrical market. Only generator active power output is rescheduled. The algorithm is tested on the modified IEEE 30-bus system.

Keywords— *congestion management, Cuckoo Search Algorithm, Independent System Operator, network congestion, power flow, rescheduling, transmission systems.*

I. INTRODUCTION

Energy is the basic necessity for economic development of a country. It exists in different forms in nature, the most refined form being electrical energy. A power system consists of generation stations, power substations, transmission lines, distribution system and load/consumption [1].
Recent global trend has been on restructuring of the Electricity Supply Industry (ESI). There has been a rapid shift from Vertically Integrated Utilities to a liberalized power market with unbundling of the generation, transmission and distribution functions [2].
In the liberalized power systems, the transmission network is a key component in enabling operation of a competitive market. It is a regulated natural monopoly which should allow open access to facilitate a competitive environment in power generation and retail services. The planning and operation of the transmission network still remains a challenge in the development of the liberalized power system. Transmission systems are at most times operated at or near their rated capacity. This is due to an increase in the number of transactions from the various market participants (buyers and sellers) as they try to make full utilization of the existing resources. In addition, there is a slow rate in construction of new transmission lines due to economic and environmental factors. As a result, congestion management is a top transmission network management problem that the ISO has to address frequently to avoid a power system crisis [3] [4].
Congestion is the restriction of transfers between different system nodes or regions in an electrical network which occurs when the transmission system is unable to accommodate all of their desired transactions due to violation of thermal limits of transmission lines. It may result from a change in energy demand or due to contingencies such as generator failure or transmission line outage [3]. Moreover, since the investment in and location of new generation plants is dictated by market forces and may not be well coordinated with transmission expansion planning, network congestion may result [6]. It is a serious economic and reliability concern in the open electricity market [5]. Thus, congestion management is a fundamental transmission system management problem in the restructured power market if secure, efficient and non-discriminatory access to the transmission system is to be maintained. Some of the remedial actions to handle congestion include outaging of congested lines, use of

Irungu G. Wangunyu, Department of Electrical and Electronic Engineering, JKUAT (corresponding author phone: +254710176121; e-mail: jerriirungu@gmail.com).

David K. Murage, Department of Electrical and Electronic Engineering, JKUAT (e-mail: davidkinyua76@gmail.com), Peter K. Kihato, Department of Electrical and Electronic Engineering, JKUAT (e-mail:)

Flexoble AC Transmission System (FACTS) devices, generator rescheduling and load shedding [7].

II. LITERATURE REVIEW

Bompard et al [6] provide a comparative analysis of congestion management schemes used in England and Wales, Norway, Sweden, PJM and California markets in the United States of America. From this work, the researchers developed a unified framework for comparing the performance of the various methods used in these power markets.

Yamina and Shahidehpour [8] developed a coordination process between Generating Companies (GENCOs) and ISO for congestion management before real-time operation using a two-stage Security-constrained Price-based Unit Commitment (SPUC). First, GENCOs apply Price-Based Unit Commitment without transmission security constraints and submit profit maximization bids to the ISO. The ISO checks for violation of transmission line limits under normal operating conditions and during contingency. If there are transmission line limit violations, a rescheduling signal is sent to the GENCOs.

Kumar et al proposed a zonal congestion management approach in [9] while in [10], they proposed a cluster-based congestion management approach. In [9], congestion zones have been determined using active and reactive power flow sensitivity indices. Generators having strongest and non-uniform sensitivity indices are grouped into most sensitive zones and their generation is rescheduled. In [10], clusters of different types are formed based on congestion distribution factors whereby type 1 clusters constitute users with the strongest and non-uniform Transmission Congestion Distribution Factors (TCDFs). In both, network analysis is performed using AC load flow method. A comparison is done between DC load flow method and AC load flow method. The findings show that the computational time while using AC load flow method is close to that of DC load flow method. Also, the AC load flow method is more accurate and it provides lower congestion costs compared to DC load flow method.

Sarwar and Siddiqui [11] achieve zonal congestion management by utilizing Locational Marginal Pricing. A power network is split into zones based on the difference in Locational Marginal Price of the buses across a transmission line. The most congestive zone, designated as “Zone 1", is identified as the one which groups buses with high and non-uniform Locational Marginal Price (LMP). A second zone, denoted as “Zone 2", is less prone to congestion and constitutes of buses with low and uniform LMP differences. A comparative study of Distributed Generator (DG) placement in Zone 1 and Zone 2 indicate that DG placement in the more sensitive Zone 1 is more effective in congestion alleviation.

Yesuratnam and Tukkaram [12] introduced a concept for mitigating transmission overload by real power rescheduling based on Relative Electrical Distance (RED). The RED is the relative location of load buses. The proposed approach alienates congestion by re-scheduling of generators by involving minimal cost in the rescheduling process. However, generators with the same Relative Electrical Distance have an equal power contribution to the congested line. The cost would not be optimized if the generators with equal RED have different incremental or decremental costs.

Sudipta and Singh [13] proposed a methodology for congestion management in a pool market by optimal rescheduling using Particle Swarm Optimization (PSO). Optimal generator rescheduling was done based on generator sensitivity to congested line's power flow.

Balaraman and Kamaraj [14] present a method for congestion management by real power rescheduling using PSO in a pool-based electrical market. Line overload due to unexpected line outage and sudden load variations are considered. Numerical results obtained using PSO are presented and compared with Simulated Annealing (SA) and Random Search Method (RSM). The experimental results show that PSO is capable of obtaining higher quality results than SA and RSM.

Ravindrakumar and Chandramohan [15] propose use of Non-dominated Sorting Genetic Algorithm II (NSGA-II) for line overload alleviation by generator rescheduling in a pool-based electricity market. The aim of the research was to reduce the rescheduling cost. NSGA II is a modified version of Non-dominated Sorting Genetic Algorithm (NSGA). NSGA is a popular non-domination based genetic algorithm for multi-objective optimization. It is a very effective algorithm but has been criticized for its computational complexity, lack of elitism and a need for prior selection of a sharing parameter value. NSGA II has a better sorting algorithm, incorporates elitism and no sharing parameter needs to be chosen a priori. Contingency cases considered in this work were line overload due to unexpected line outage and sudden load variations.

There has also been research focusing on the application of FACTS devices such as Thyristor-Controlled Series Compensator (TCSC), Thyristor Controlled Phase Angle Regulator (TCPAR) and Static Var Compensator (SVC) to solve congestion problems. Mwanza [16], reports that inclusion of FACTS devices greatly reduces amount of redispatched power in a pool-based market. This has the advantage of maintaining the optimal operating point close to that obtained from market settlement. Also, there is a reduction in the congestion cost incurred by the ISO. Rajalakshmi et al [17] present a method for optimal location of TCSC based on real power performance index and reduction of total system VAR power loss. Bhattacharyya et al [18] implemented a fuzzy-DE approach for optimal placement of TCSC and SVC with the objective of minimizing transmission loss while maintaining lowest operating cost. TCSC's were placed in lines with very high reactive power flow while the placement of SVC's was determined by the fuzzy membership of loss sensitivity in the weaker nodes. The results obtained showed a better performance of Fuzzy-DE compared to SDE (Simple Differential Evolution). However, FACTS devices involve high capital cost and a long payback period. Their use is only plausible if they are economically justifiable alternatives to power system reinforcements [7].

Cuckoo Search Algorithm is a nature-inspired metaheuristic approach developed in 2009 by Xin-She Yang and Suash Deb. It has been applied to solve problems in engineering, pattern recognition, job scheduling, Object-Oriented software testing and in wireless sensor networks [19].

III. MODELLING OF THE ENERGY MANAGEMENT PROBLEM

Congestion management aims at minimizing network congestion while satisfying network constraints. The problem is formulated as [14];

$$minimize\ C_c = \sum_{j=1}^{N_g}(C_k\Delta P_{Gi}^{+} + D_k\Delta P_{Gi}^{-})\ \$/h \quad (2.1)$$

The optimization problem is subject to equality and inequality constraints as stated in the next two subsections.

A. Equality constraints

$$P_i = |V_i|\sum_{j=1}^{n}|V_j||Y_{ij}|\,cos(\theta_{ij} + \delta_j - \delta_i) \quad (2.2)$$

$$Q_i = -|V_i|\sum_{j=1}^{n}|V_j||Y_{ij}|\,sin(\theta_{ij} + \delta_j - \delta_i) \quad (2.3)$$

$$P_{Gi} - P_{Di} - P_i = 0 \quad (2.4)$$

$$Q_{Gi} - Q_{Di} - Q_i = 0 \quad (2.5)$$

B. Inequality constraints

These represent the operating and physical limits of all the generators, transformers and transmission lines, as stated in equations *2.6-2.9*;

$$P_{Gi}^{min} \le P_{Gi} \le P_{Gi}^{max} \quad (2.6)$$

$$Q_{Gi}^{min} \le Q_{Gi} \le Q_{Gi}^{max} \quad (2.7)$$

$$V^{min} \le V \le V^{max} \quad (2.8)$$

$$P_{ij} \le P_{ij}^{max} \quad (2.9)$$

Where;

P_{Gi} and P_{Di} are the active power generation and demand at bus *i* respectively,

Q_{Gi} and Q_{Di} are the reactive power generation and demand at bus *i* respectively,

N_g is the number of generators in the power system.

P_{Gi}^{min} and P_{Gi}^{max} are the minimum and maximum active power generation limits,

Q_{Gi}^{min} and Q_{Gi}^{max} are the minimum and maximum reactive power generation limits,

V^{min} and V^{max} are the bus minimum and maximum voltage limits,

P_{ij}^{max} is the maximum line flow limit (MW) of line l,

Equations *2.2* and *2.3* represent the power flow equations for a given bus *i* while equations *2.4* and *2.5* represent the power balance constraints.

IV. CUCKOO SEARCH ALGORITHM

Cuckoo Search Algorithm (CSA) is a nature-inspired metaheuristic algorithm based on the obligate brood parasitism of some cuckoo species such as *Tapera* in combination with the Levy flight behaviour of some birds and fruit flies. In brood parasitism, the birds lay eggs in nests of other host birds, often of other species. Some of the host birds can engage in direct conflict with the intruders. In other instances, if the host bird discovers the alien eggs, it may throw them away or abandon the nest. To reduce the probability of the eggs being abandoned by the host bird, some cuckoo species such as *Tapera* have evolved and are specialized in the mimicry in color and pattern of the eggs of some of the host species. The parasitic cuckoo often chooses a nest where the host bird has just laid its eggs. The cuckoo eggs have slightly shorter incubation period than the host's eggs. Upon hatching, the first instinctive action of the cuckoo chick is to blindly propel the other eggs out of the nest. This increases its share of food provided by the host. The chicks have rapid nestling growth and can also mimic the call of host chick, increasing its feeding opportunity [20].

In this algorithm, the eggs in a nest represent a pool of candidate solutions of an optimization problem, while the cuckoo egg represents a new coming solution. The aim is to use these new, and potentially better solutions associated with the cuckoo eggs to replace the current solution associated with the eggs in the nest. It is based on three rules [21].

1. Each cuckoo lays one egg at a time and dumps it in a randomly chosen nest.
2. The best nests with high quality of eggs (solutions) will be carried over to the next generations.
3. The number of available host nests is fixed, and a host can discover an alien egg with a probability p_a [0; 1]. The host bird can either throw the egg away or abandon the nest so as to build a completely new nest in a new location. This rule can be approximated by a fraction p_a of the *n* nests being replaced by new nests with new random solutions at new locations.

The algorithm uses a balanced combination of local random walk and global exploration. It can be summarized in a pseudo-code as;

begin
Objective function f(x), x = (x_1; x_2 .....x_d);
Initialize a population of n host nests x_i (i=1, 2..... n);
while *(t< maximum generation) or (stop criterion);*
Get a cuckoo i randomly;
Generate a new solution by Levy flight;
Evaluate its fitness, F_i;
Randomly choose a nest j among the n nests;
if *(F_i > F_j),*
Replace j with the new solution;
endif
Abandon a fraction pa of worse nests;
Build new nests at new locations by Levy flights;
Keep the best solutions;
Rank the solutions and find the current best;
end while
Output the results
end

V. METHODOLOGY

This work explores the capability of Cuckoo Search Algorithm in solving the congestion management problem. In performing the analysis, modified IEEE 30-bus system was used as the test-bed. Outage of line 1-2 was the considered contingency case. Only generator active power rescheduling was performed. The set of control variables include changes in generator active power output, excluding the slack bus.

A. Modified IEEE 30-bus system

This system has a total active power demand of 283.4 MW and reactive power demand of 126.2 MVAr. It consists of 6 generator buses, 24 load buses and 41 transmission lines [14].

B. Computational procedure of CSA for CM

1. Bus data, line data, price bids and generator parameters were input and a load flow was run for the base case scenario.
2. A contingency was created by introducing an outage on line 1-2.
3. Load flow was run while satisfying the equality constraints. Hence, excess power flow was determined.
4. Initial Cuckoo population was generated using equation (2.6), which is the amount of rescheduling required by the generators to manage congestion (randomly within the limits).
5. For each generated Cuckoo population, load flow was performed and, hence, the fitness function was evaluated and the best solution was identified.
6. New solutions were generated using Levy flight.
7. The fitness of each new solution obtained was evaluated. Any two nests were randomly chosen and their fitness values were compared. The cuckoo with better fitness value was accepted while the other was rejected.
8. Steps 6-7 were repeated until the congestion is eliminated.

VI. RESULTS AND DISCUSSION

CSA for congestion management was implemented using MATLAB (version R2014a) software. Simulations were carried out on modified IEEE 30-bus system. Congestion was created by outage of line-1 connected between bus-1 and bus-2. Generator rescheduling cost for the simulated case was calculated and compared with results reported in [14].

Due to outage of line-1, congestion occurs on line-2 and line-4, connected between buses 1-7 and 7-8, respectively. From load flow results, due to the outage, the power flow in these lines become 147.263 MW and 135.960 MW respectively, against the line flow limit of 130 MW for both lines. Net power violation was found to be 23.223 MW. Results obtained after performing generator rescheduling are as tabulated in *Table 1*. *Table 1* also includes results obtained from RSM, SA and PSO techniques reported in [32]. From Table 1, it is evident that CSA gives the lowest rescheduling cost compared to RSM, SA and PSO. CSA gives a rescheduling cost of 487.5818 $/h. The total system loss after congestion management decreases from 15.823 MW to 12.922 MW. A comparative pictorial representation of active power rescheduling and congestion cost offered by CSA, PSO, RSM and SA are shown in *Figure 1* and *Figure 2*, respectively. *Figure 3* shows the bus voltage profile obtained after CM using CSA. It is observed that the voltage magnitude at each bus is within limit.

Table 1: Comparison of results obtained.

Parameters	**Method**			
	CSA	**PSO**	**RSM**	**SA**
Total congestion cost ($/h)	487.5818	538.95	716.25	719.861
Power flow after CM(MW)				
Line 1-7	129.2950	129.9700	129.7800	129.5100
Line 7-8	120.2600	120.7800	120.6000	120.3500
ΔP_{G1} (MW)	-9.5090	-8.6123	-8.8086	-9.0763
ΔP_{G2} (MW)	15.0668	10.4059	2.6473	3.1332
ΔP_{G3} (MW)	0.0000	3.0344	2.9537	3.2345
ΔP_{G4} (MW)	0.0001	0.0170	3.0632	2.9681
ΔP_{G5} (MW)	0.0002	0.8547	2.9136	2.9540
ΔP_{G6} (MW)	0.0001	-0.0122	2.9522	2.4437
Total generation rescheduled (MW)	24.5762	22.9360	23.3390	23.8090

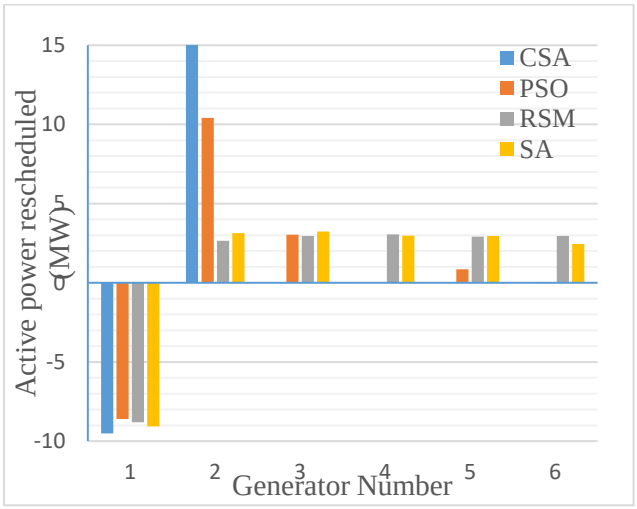


Figure 1: Comparative active power rescheduling of generators.

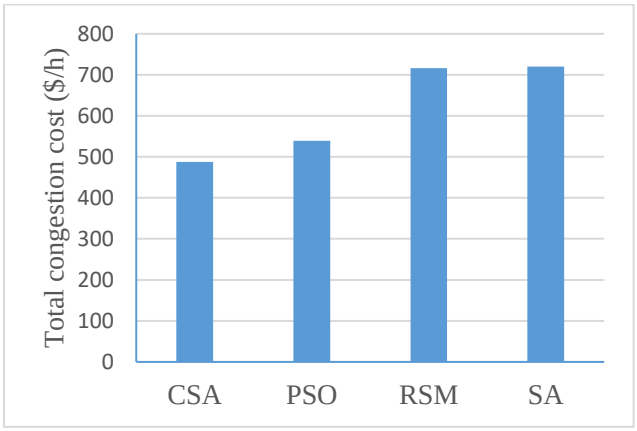


Figure 2: Congestion cost obtained from different algorithms.

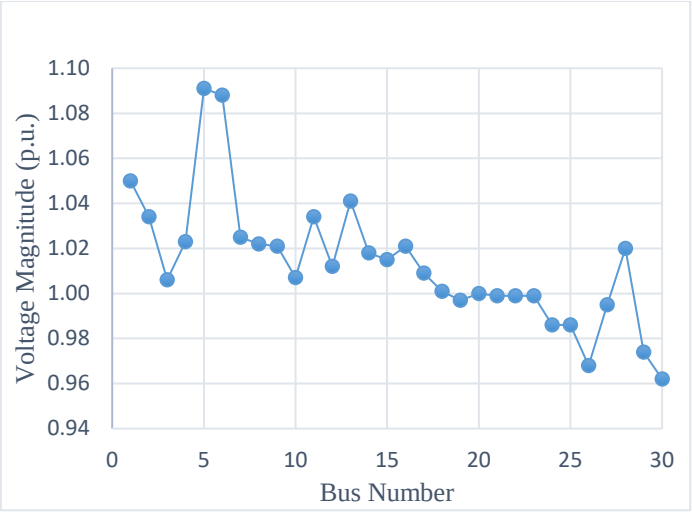


Figure 3: Voltage profile after Congestion Management

VII. Conclusion

This paper presents a generator rescheduling-based approach for power system Congestion Management. A line outage is the only contingency case considered. Congestion in the overloaded lines is managed by generator rescheduling while minimizing the rescheduling cost. The comparative analysis of the obtained results indicates that the proposed CSA-based algorithm gives lower rescheduling cost compared to Particle Swarm Optimization, Simulated Annealing and Random Search Method. In addition, the total system loss has also been reduced after performing Congestion Management.

References

[1] D. Das, *Electrical Power Systems*, New Delhi: New Age International (P) Limited, 2006.

[2] J. Tooraj, N. Rabindra and G. R. Timilsina, "A quarter century yet to come of age: a survey of power sector reforms in developing countries," Policy Research working paper; no. WPS 7330. Washington, D.C. : World Bank Group, 2015.

[3] S. Chanda and A. De, "Application of particle swarm optimisation for relieving congestion in deregulated power system," *Recent Advances in Intelligent Computational Systems (RAICS)*, pp. 837-840, 2011.

[4] A. Pillay, S.P. Karthikeyan and D.P. Kothari, "Congestion management in power systems," *Electrical Power and Energy Systems*, vol. 70, pp. 83-90, 2015.

[5] R. D. Christie, B. F. Wollenberg and I. Wangensteen, "Transmission Management in the Deregulated Environment," in *Proceedings of the IEEE*, 2000.

[6] E. Bompard, P. Correia, G. Gross and M. Amelin, "Congestion-Management Schemes: A comparative analysis under a unified framework," *IEEE Transactions on Power Systems*, vol. 18, no. 1, pp. 346-352, 2003.

[7] A. Singla, K. Singh and V.K. Yadav, "Transmission congestion management in deregulated environment: A bibliographic survey," in *Proceedings of the IEEE International conference on recent advances and innovations in engineering*, Jaipur, India, 2014.

[8] H. Y. Yamina and M. Shahidehpour, "Congestion management coordination in the deregulated power market," *Electric Power Systems Research*, vol. 65, no. 2, pp. 119-127, 2003.

[9] A. Kumar, S. C. Srivastava and S. N. Singh, "A zonal congestion management approach using real and reactive power rescheduling," *Electric Power Systems Research*, vol. 19, no. 1, pp. 554-562, 2004.

[10] A. Kumar, S. C. Srivastava and S. N. Singh, "A zonal congestion management approach using ac transmission congestion distribution factors," *Electric Power System Research*, vol. 72, pp. 85-93, 2004.

[11] M. Sarwar and A. S. Siddiqui, "Zonal congestion management based on locational marginal price in deregulated electricity market," in *2015 Annual IEEE India Conference (INDICON)*, New Delhi, India, 2015.

[12] G. Yesuratnam and D. Thukaram, "Congestion management in open access based on relative electrical distance using voltage stability criterion," *Electrical Power Systems Research*, vol. 77, pp. 1608-1618, 2007.

[13] S. Dutta and S.P. Singh, "\Optimal Rescheduling of Generators for Congestion Management Based on Particle Swarm Optimization," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 1560-1569, 2008.

[14] S. Balaraman and N. Kamaraj, "Transmission congestion management using Particle Swarm Optimization," *Journal of Electrical Systems*, vol. 7, no. 1, pp. 54-70, 2011.

[15] K.B. Ravindrakumar and S. Chandramohan, "NSGA II based Congestion Management in Deregulated Power Systems," *Middle-East Journal of Scientific Research*, vol. 24, no. 4, pp. 1188-1193, 2016.

[16] K. Mwanza and Y. Shi, "Congestion Management: Re-dispatch and Application of FACTS," 2006.

[17] L.Rajalakshmi, M.V.Suganyadevi and S. Parameswari, "Congestion Management in Deregulated Power System by Locating Series FACTS Devices," *International Journal of Computer Applications*, vol. 13, no. 8, pp. 19-22, 2011.

[18] B. Bhattacharyya, V. K. Gupta and S. Kumar, "Fuzzy-DE approach for the optimal placement of FACTS devices to relief Congestion in a power system," in *Proceedings of The 2014 International Conference on Control, Instrumentation, Energy and Communication (CIEC)*, Calcutta, 2014.

[19] A.B Mohamad, A.M Zain and N.E.N Bazin , "Cuckoo Search Algorithm for Optimization Problems- A Literature Review and Applications," *Applied Artificial Intelligence: An International Journal*, vol. 25, no. 5, pp. 419-448, 2014.

[20] A. H. Gandomi, X. Yang and A. H. Alavi, "Cuckoo search algorithm: a metaheuristic approach to solve structural optimization problems," *Engineering with Computers*, vol. 29, pp. 17-35, 2013.

[21] A. Galvez, A. Inglesias and L. Cabellos, "Cuckoo Search with Levy Flights for Weighted Bayesian Energy Functional Optimization in Global-Support Curve Data Fitting," *The Scientific World Journal*, pp. 1-11, 2014.

Power System Restructuring: A review of the progress in various parts of the world

Irungu G. Wangunyu[a], David K. Murage[b] and Peter K. Kihato[c]
[a,b,c]Department of Electrical and Electronic Engineering, Jomo Kenyatta University of Agriculture and Technology, Kenya

Abstract— Globally, the electrical power supply industry has been and is still undergoing change in the mode of operation from Vertically Integrated Utilities (VIUs) to liberalized power markets, with unbundling of the generation, transmission and distribution functions. These reforms have been driven by various internal and external factors. The internal drivers for electrical industry liberalization vary depending on the level of development of a country.
In mature economies, restructuring has been propelled by the need to provide electricity at lower prices and to offer customers a greater choice in the retail market. In developing countries, the power industry needs to meet the fast growing electrical energy demand induced by blooming economies. The overall expected outcomes are improved efficiency, improved service reliability and introduction of reflective pricing, hence, a reduction in the price-cost gap of electrical energy. These benefits are expected to stimulate economic development.
This paper reviews the drivers of the reforms; the proposed standard model of implementing the reforms; merits and demerits of the various options which chart the evolution of the Electricity Supply Industry (ESI) from a regulated monopoly to full competition. In addition it discusses the various fully restructured power markets in the world; the reforms, so far, in Sub-Saharan Africa, South Asia and Latin America; their impacts and lessons learned.

Keywords— *Vertically Integrated Utilities, Liberalized Power Markets, Restructuring, Electricity Supply Industry, Standard Reforms Model, Regulated monopoly, Power markets.*

I. INTRODUCTION

Adequate, reliable and competitively priced electricity is essential for modernization, domestic growth and international competitiveness [1]. Reliable electrical power supply has become increasingly important since industrial, commercial and domestic consumers are relying on electronic devices to perform a wide range of tasks. In addition, a reliable power supply facilitates the provision of modern basic services such as healthcare and communication [2].
Historically, the ESI has been a monolithic structure with a single entity owning generation, transmission and distribution infrastructure in addition to performing all system operations. In many countries, especially developing countries, the electric utility was owned by the state or local government while in other countries, such as in the United States of America, it was an investor-owned monopoly. The VIUs were a natural monopoly since only the national or local electric utility was permitted to produce, transmit, distribute and sell electric power within its service territory. The governments guaranteed regulated rates that would provide the electric utilities with a profit margin above their operating cost. At most times, this was replaced by economic arrangements such as government ownership and subsidies. This operation structure was inadequate in providing incentives for innovation since the utilities had little motivation to use new ideas and technologies to lower costs under a regulated rate of return framework. Lack of competition also gave the utilities little incentive to improve standard of service to the power consumers [3].

II. DRIVERS OF REFORMS IN THE ESI

Since Chile pioneered a radical restructuring and privatization in 1982, the shift towards liberalized power markets has been a global phenomenon [1]. The reforms have been driven by new requirements for the energy industry such as reducing electrical energy cost, improving operational efficiency and maintaining sustainable long-term development of the power system infrastructure. Liberalization will enable to satisfy these requirements by creating a competitive environment for electricity trading. A competitive power industry will provide rewards to risk takers and encourage the use of new technologies. This shall drive electricity price closer to the marginal cost of generation, generating companies will optimize their production cost and consumers could get a cheaper electrical energy cost with a choice of the supplier, resulting in improvement of social welfare [3] [4].
Specific reasons for adoption of Electrical Supply Industry (ESI) restructuring vary depending in the level of economic development of a country. In developing countries, inadequate public sector financial resources to meet growing demand; institutional inefficiency; burden of energy subsidies; low service quality; high energy losses and capacity shortage needed to be addressed. In the mature industrial economies, the pressure for change grew with; the emergence of excess

Irungu G. Wangunyu, Department of Electrical and Electronic Engineering, JKUAT (corresponding author phone: +254710176121; e-mail: jerriirungu@gmail.com).

David K. Murage, Department of Electrical and Electronic Engineering, JKUAT (e-mail: davidkinyua76@gmail.com), Peter K. Kihato, Department of Electrical and Electronic Engineering, JKUAT (e-mail:)

capacity, disillusionment with expensive and capital-intensive generation projects precipitated by the oil crisis of the 1970s, and a need offer customers a great choice in the retail market [5].

Other factors that have contributed to the restructuring are technology advances; political faith on the forces of market, competition and privatization; the rise of environmentalism; the shortage of public resources for investment in developing countries; and pressures from international donor agencies. Moreover, advances in gas turbine technology has resulted in more efficient small turbines and generators which could match the efficiency of very large units. This has led to a change in generation economies of scale [3].

III. Options for Restructuring Electricity Markets

An ideal fully restructured power industry should have competitive generation and retail markets along with the regulated transmission and distribution networks. A general representation of the restructuring process is as illustrated in *Fig.1* [6].

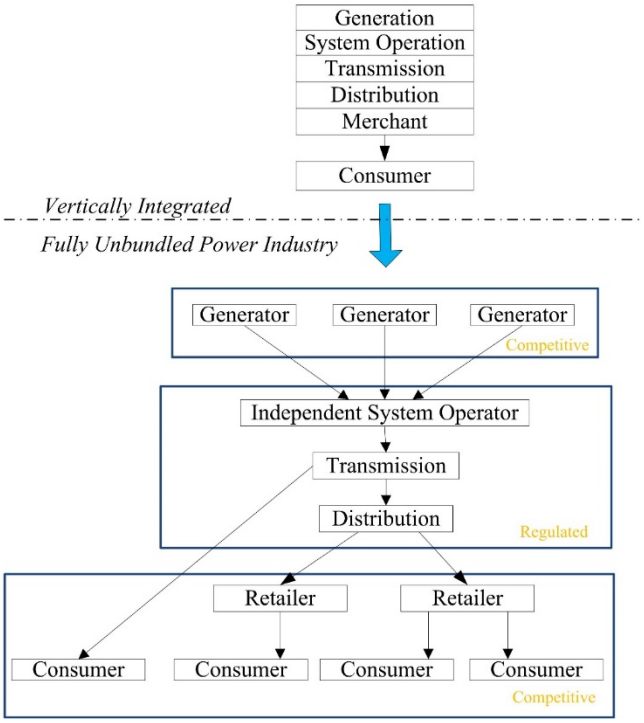


Fig.1: Evolution of the electrical power industry [6].

There are four models, as proposed by Hunt and Shuttleworth, which chart the evolution of the ESI from a regulated monopoly to full competition [7];

Model 1: monopoly at all levels. This model corresponds to the traditional monopoly utility structure. There is no competition at the power generation level and the consumer has no choice in selecting the supplier. A single company has the monopoly of producing electricity and delivering it over the transmission and distribution networks to final consumers.

Model 2: purchasing agency/single buyer. This is the first possible step towards introduction of competition in the ESI. The model allows a single buyer to choose from a number of generators to encourage competition in generation. This model is perceived to be a reasonable second-best solution in countries where Model 3 would not work [8]. It could take two forms as illustrated in *Fig. 2*. In *Fig. 2(a)*, there are IPPs connected to the network and sell their output to the vertical utility, which acts as a purchasing agent. A further evolution of this model is as illustrated in *Fig. 2(b)* whereby the utility no longer owns any generation capacity and purchases all its energy from IPPs. There is also a disaggregation of distribution and retail activities. DISCOs purchase the energy consumed by their customers from the wholesale purchasing agency.

The rates set by the purchasing agency must be regulated to avoid it exercising market power. This model does not discover a cost-reflective price in the same way that retail competition does. However, it has the advantage of introducing some competition between generators without the expense of setting up a competitive market.

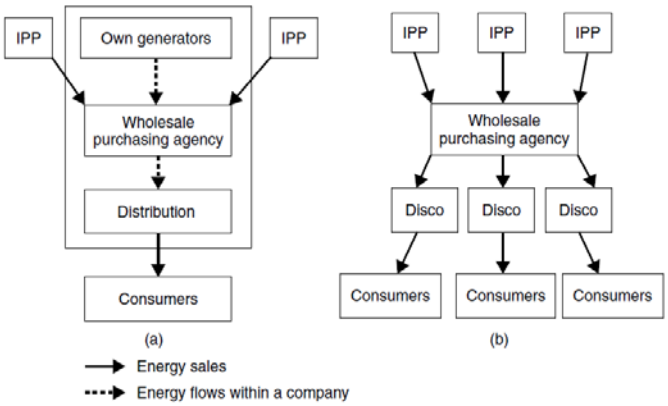


Fig. 2: Purchasing agency model of electricity market; (a) integrated version (b) disaggregated version [7].

Model 3: wholesale competition. In this model, DISCOs purchase the electrical energy consumed by their customers directly from generating companies. The transactions take place in a wholesale electricity market. In addition, large consumers are also allowed to make direct energy purchases from the wholesale market. There is open access to transmission grid. The wholesale market can take the form of a pool or bilateral transactions.

At the wholesale level, the only functions that remain centralized are the operation of the spot market and the operation of the transmission network. Each DISCO still has a monopoly over final consumers located in its service territory. There is considerably more competition between GENCOs since the wholesale price is determined by the interplay of supply and demand. However, the retail price of electrical energy must remain regulated since small consumers cannot choose a competing supplier if they feel that the price is too high.

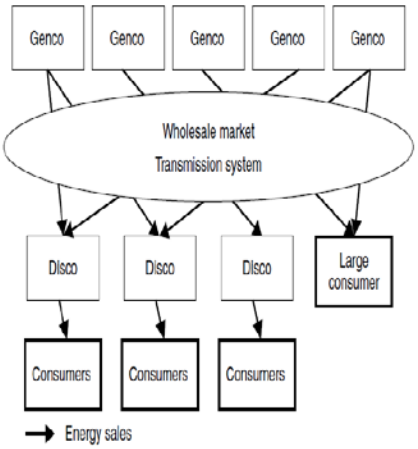


Fig. 3: Wholesale competition [7].

Model 4: retail competition. In this model, all consumers can choose their supplier. There is open access to transmission and distribution networks. The distribution is separate from the retail activity, which is fully competitive. Because of the transaction costs, only the largest consumers choose to purchase energy directly on the wholesale market. Most small and medium consumers purchase power from retailers, who in turn buy it in the wholesale market. This model makes the most competitive forces by bringing all final consumers into the market. However, it also greatly increases transaction costs due to requiring more complex trade arrangements and metering.

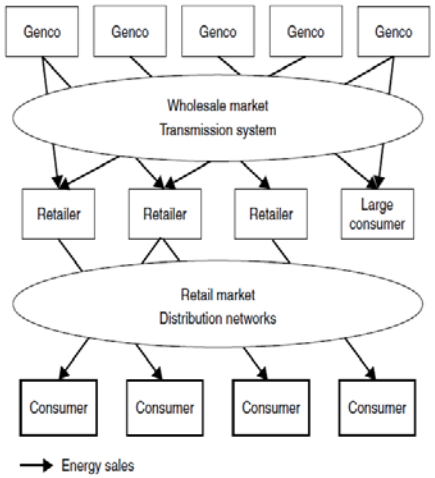


Fig. 4: Retail competition [7].

IV. STANDARD MODEL OF POWER-SECTOR REFORM/STANDARD TEXTBOOK MODEL

The standard model is also referred to as the standard textbook model. It was first applied in the Chilean power sector in 1982 and also became the reference model for reforms in other energy sectors. It prescribes the basic architecture for electricity restructuring and involves the following reform sequence and steps [9]:

1. *Corporatization:* Transforming the power utility into a separate legal entity, with all associated rights and obligations including governance structures, managing budgets, borrowing, procurement, labour employment, payment of taxes and dividends.
2. *Commercialization:* introduction of cost-recovery pricing, improvements in metering, billing and collections, adopting internationally recognized accounting practices as well as accounting for all subsidies.
3. *Requisite legislation:* Passing legislation that provides a legal mandate for restructuring, as well as the legal framework to allow private/foreign participation/ownership in the sector.
4. *Establishment of an independent regulator:* aims to introduce efficiency, transparency and fairness in the management of the sector, specifically to prevent anticompetitive activity, encourage appropriate investment and protect consumers.
5. *Independent Power Producers (IPPs):* introduce new (private) investment in generation, with long-term Power Purchase Agreements (PPA).
6. *Vertical and horizontal restructuring:* to separate potentially competitive generation and retail activities from the natural monopoly segments of transmission and distribution and thus facilitate competitive entry and mitigate market power.
7. *Divestiture of generation assets:* divesting state ownership (in part or in full) of generation assets to the private sector.
8. *Divestiture of distribution assets:* divesting state ownership (in part or in full) of distribution assets to the private sector.
9. *Competition:* introduction of wholesale and retail markets.

V. EXAMPLES OF FULLY RESTRUCTURED POWER MARKETS

A. United Kingdom

- The old England & Wales Pool

After restructuring, the vertically integrated Central Electricity Generating Board (CEGB) in England and Wales had been split into three generating companies: National Power, Powergen and Nuclear Electric [10]; one transmission company: the National Grid Company and twelve regional electricity companies (RECs) which were regarded as local distribution systems [1].The National Grid Company had two roles: transmission asset owner (TO) and Independent System Operator (ISO) [10] [11]. An electricity pool market had been established which arranged and managed electricity trades in England and Wales with wholesale market mechanism. *Fig. 5* shows the structure of electricity industry at privatization [adapted from [10]].

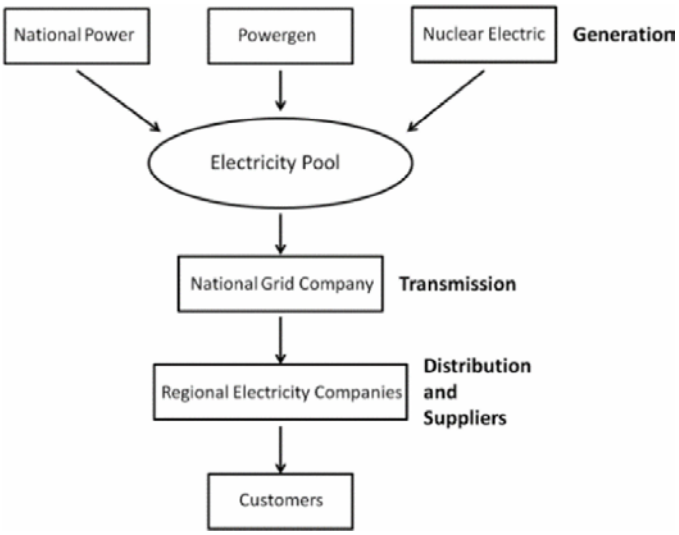


Fig. 5: Structure of Electricity Industry in England and Wales.

The old England & Wales Pool was a typical mandatory pool market; every energy trade had to be transacted only inside the pool market [12]. The SO would accept bids from the cheapest price to higher price until the forecasted demand was satisfied. Then, the SO would sort out a bids list containing names of the generation companies that had been chosen to generate electricity on the following day. Those generation companies were called "in merit" generation companies which means their bids have been accepted by the SO and those who have not been accepted by the SO were called "out of merit" generation companies [12]. At last, the SO would set prices for all pool market participants.

System Marginal Price (SMP) is the price bid by the most expensive generation company in the bids list [1]. In the pool market, the SO would pay for the Pool Purchase Price (PPP) to all the in merit generation companies. The Pool Purchase Price (PPP) is the System Marginal Price (SMP) plus Capacity Payment (CP) which is defined with Loss of Load Probability (LOLP) and Value of Lost Load (VOLL) [13]. On the other side, suppliers would pay for the price called Pool Selling Price (PSP) to purchase the electrical power from the pool [13]. The PSP is the PPP plus the Uplifts. The following equations defined the relationships between those prices [11]

$$PPP = SMP + CP \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (1)$$
$$CP = LOLP \times MAX(0, VOLL - SMP) \dots\dots\dots\dots (2)$$
$$PSP = PPP + Uplifts \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3)$$

CP is an incentive which is used to reward the generation companies who declare that their capacity is available regardless of whether they are required to or not. LOLP is the probability that electricity power capacity is unable to support the actual demand. LOLP is predicted in each half hour.

The pool structure is regarded as an innovation for energy industry restructuring [10]. However, electricity price decrease was slower than generator cost reduction and customers were unable to access to the deregulation benefits. The main reason is that, at most times, the three dominant generation companies could exercise their market power and determine price [14].

- *The British Electricity Trading and Transmission Arrangements (BETTA)*

On 27th March 2001, the New Electricity Trading Arrangements (NETA) was introduced to replace the pool with the concept that markets participants have rights to transact electricity power by bilateral trading [10]. On 1st April 2005, with the joining of Scottish network, the New Electricity Trading Arrangements (NETA) has been extended as the British Electricity Trading and Transmission Arrangements (BETTA) [10]. In the BETTA, electricity participants are able to trade by bilateral contracts. The bulk of electricity power is traded by Forward/Future contracts and short-term power exchange [15]. The Forward/Future contracts could be signed ahead of days, months or even years before actual power delivery [13]. The short-term power exchange is within 24 hours before electricity delivery which offers market participants an opportunity to adjust their contract details. Market participants are required to notify their predicted physical information in each half hour period, for seller is the planned generation output and for buyer is the metered demand, by 11 a.m. one day ahead. This information is called initial physical notifications (IPNs) [10]. Then market participants hand in their final physical notifications (FPNs) to the SO before the gate closure, usually 3.5 hours before delivery [10]. Figure 2 show the structure of BETTA [5] and figure 3 shows the diagrammatic representation of BETTA time frame process [12].

Apart from physical data, participants have an opportunity to voluntarily submit their offers and bids to the SO to take part in balancing mechanism [15]. The SO collects offers and bids then matches the balance between supply and demand so as to ensure system security. After offers and bids have been accepted by the SO, market participants will be paid or charged according to their contracted prices. The SO will punish the participant who violates the contract.

Because more price information is available, bilateral market provides more liquidity so that electricity price decreases remarkably [15]. However, in order to avoid penalty from the SO, generators might increase output whilst suppliers might reduce demand which will make inefficient elements arise. Another problem is that renewable generation is unable to guarantee the production accurately so it will be easier penalized than other generations [15].

B. Nordic Countries

The Nordic electricity market consists of four markets from four countries: Sweden, Norway, Finland and Demark [16]. The Nord Pool consists of two physical markets and several financial markets [17]. Figure 4 shows Nordic market's components [17]. Fig. 6 shows the diagrammatic representation of Nord pool time frame process [12].

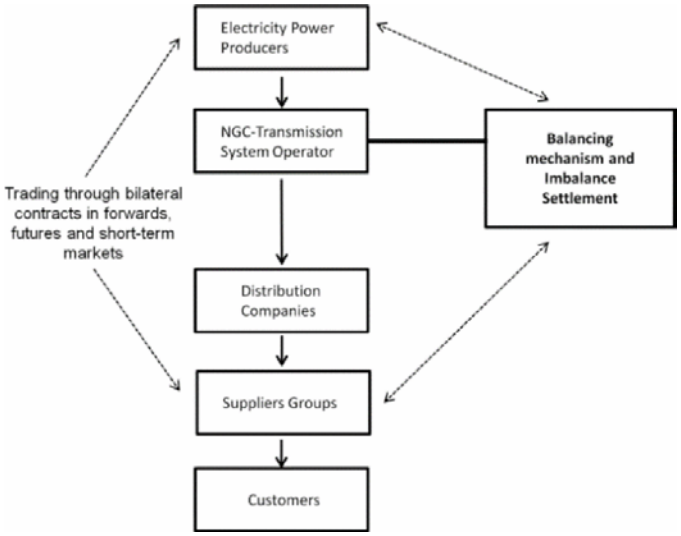


Fig. 6: The structure of BETTA.

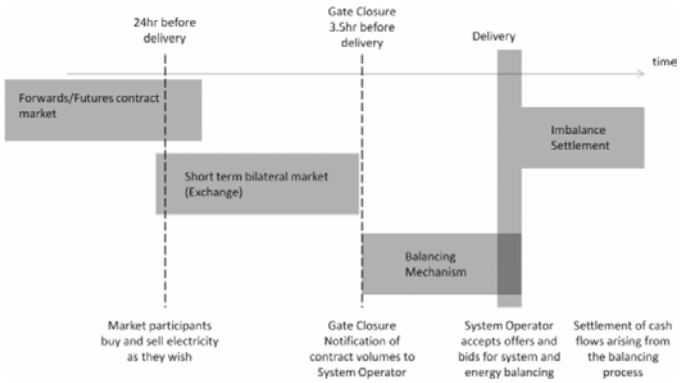


Fig. 7: Diagrammatic representation of BETTA time frame process.

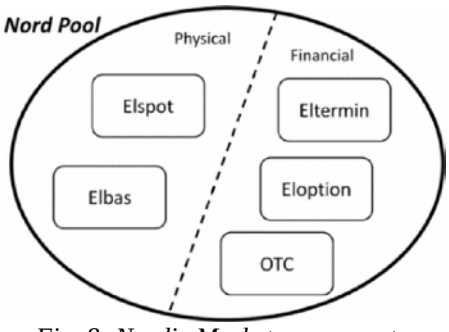


Fig. 8: Nordic Market components.

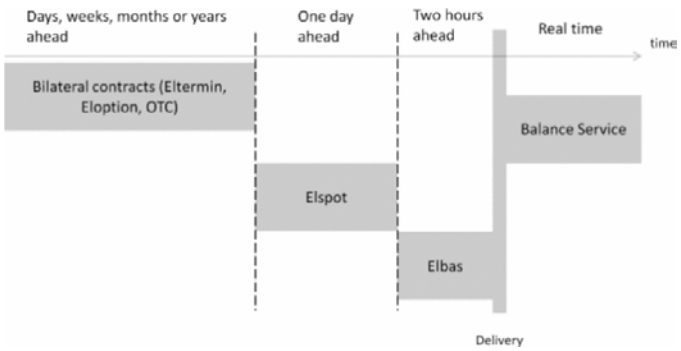


Fig. 9: Diagrammatic representation of Nord Pool time frame process.

Elspot market is referred as the main physical market which is a day-ahead market arranging hourly bilateral contracts. It determines the hourly spot price which will be regarded as the reference price both for financial and physical market. About one thirds of electricity consumption is traded in Elspot market and the rest of transactions take place through bilateral contacts [17]. Elbas is seen as a supplement of Elspot market for additional balancing services. It offers participants an opportunity to adjust their physical power position such as the price and volume of supply or demand within two hours period. There are two financial markets in Nord Pool; Eltermin and Eloption. Sellers and buyers can sign a forward and future contract in Eltermin market days, weeks, months and even up to four years ahead to hedge against the prices volatility risk. Eloption market is quite a new financial market which also provides contract services.

The Nord Pool also allows brokers to organize the over the counter (OTC) trading for market participants to make bilateral contracts [12].The Nord Pool power exchange is the key point of Nordic electricity market. System price is determined in spot market [18].When transmission network capacity is sufficient, the wholesale electricity prices in four countries are the same. Once transmission capacity shortage takes place, prices in different areas will be different [17]. Sufficient hydropower and low demand make Nordic market a nearly wonderful power system [18]. Having four electricity markets dilute the market power [18].

C. Pennsylvania-Jersey-Massachusetts (PJM) Interconnection

The PJM Interconnection is a Regional Transmission Organization (RTO) in the United States [19]. It serves the area of Pennsylvania, New Jersey, Maryland and other states in the eastern part of United States [20]. It serves 61 million of customers and has 900 participant power companies, making it the largest electricity wholesale market in the world. It has an energy market which consists of a day-ahead market and a real-time balancing market; a capacity credit market; a Financial Transmission Right (FTR) auction market and an ancillary service market which comprises of a regulation market and a pin reserve market [19]. Figure 6 shows the PJM market components [19]:

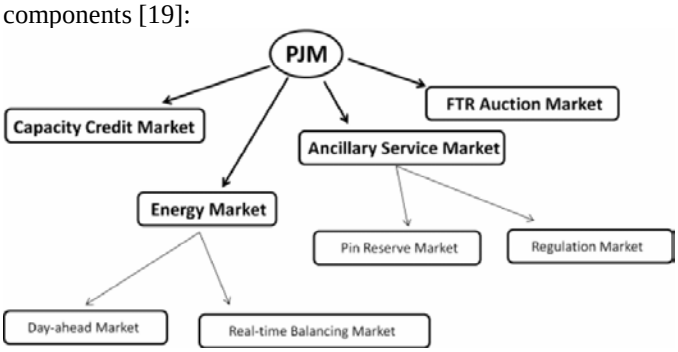


Fig. 10: PJM components.

The day-ahead market calculates the hourly clearing prices for each hour of the next operating day based on generator's offers, demand bids, virtual supply offers, virtual demand bids and bilateral transaction schedules. The balancing market calculates the clearing price in every 5 minutes period [20].

The day-ahead market settlement follows the planned hourly power quantities and day-ahead hourly electricity prices. The

balancing market settlement follows the hourly quantity deviations between planned values and real-time values. Both prices calculation are based on the mechanism of LMP. The day-ahead market allows participants to exchange electricity at binding day-head prices.

Generators who produce more than scheduled quantities will be paid by real-time LMPs. Customers who consume more than the planned power quantities are charged by real-time LMPs. If participants who exchange spot energy, their trades will be settled at real-time LMPs. Besides the day-ahead market and the real-time balancing market, PJM also provides financial instruments for participants to hedge the price variation risks. PJM can provide market liquidity and price transparency. Efficient and sufficient price signals help market participants to make correct judgement [20].

VI. Status of reforms in Sub-Saharan Africa, South Asia and Latin America

Power sector reforms have been implemented in Latin America with greater success. Most of the countries in Latin America have adopted Model 3 and considerable reforms have been carried out with adherence to the standard reform model [8]. Regulated wholesale and retail competitive arrangements are the norm. Today, most countries in this region combine competition for the market with competition in the market (mostly in balancing markets) [9].

The Sub-Saharan Africa countries are at different stages of restructuring their electricity markets from regulated monopolies to a competitive electricity market. For the most part, power systems in countries across Sub-Saharan Africa and South Asia have retained state-owned utilities, with varying degrees of unbundling, regulation, competition and private participation, mainly in the form of IPPs. The single-buyer model is dominant. Where elements of competition have been introduced, it has been competition for the market – through auctioning long-term contracts with IPPs – rather than in the market - where generators continuously compete for least-priced dispatch and sales through power exchanges and distribution companies offer competitive retail services to customers [2].

In Sub-Saharan Africa, wholesale and retail competition are entirely absent. The standard reform model has not been fully realized anywhere in Africa. The most pervasive features of power sector reform have been the corporatization and commercialization of state-owned utilities and the introduction of independent regulatory bodies. A 'hybrid market model' is emerging in many Sub-Saharan African countries. In this model, private and public investment coexist in a sector that continues to be state-dominated [9].

VII. Outcomes

In countries where electricity reforms have been pursued, there has been an improvement in the overall sector performance. Specifically, quality of service has improved as a consequence of a steady fall in power outages and a shrinking of energy losses (both technical and commercial). Also, improvement in labour productivity has been reported in some countries such as in Chile. Moreover, the reforms have better aligned prices with underlying costs. Wholesale competition has worked well in industrial countries because of excess capacity, moderate demand growth, and the availability of natural gas (which enabled the entry of gas-fired plants at modest scale and relatively low cost) [1].

VIII. Challenges

Two key challenging tasks encountered by policymakers in developing countries are [5];

(i) Formulation of electrical energy pricing schemes that balance between economic efficiency (by reversing underpricing and cross-subsidies) and social equity.
(ii) How to increase the availability of adequate, reliable, clean and competitively priced electrical power while mitigating climate change by adopting renewable energy sources.

IX. Lessons learned

The emerging international evidence suggests that: if implemented in the correct way, the standard reform model is a sound guide for successful electricity market restructuring. Significant departures from the textbook reform model are likely to lead to performance problems. Emerging consensus is that in the electricity sector [5] [21];

(i) When well designed and implemented in proper sequence, a combination of institutional reforms—vertical and horizontal restructuring, privatization, and effective regulation (particularly the application of incentive-based regulatory mechanisms)—can lead to significant improvements in several dimensions of operating performance and in a variety of country settings.
(ii) There is a strong link between good and credible regulation and the objective of securing foreign direct investment, and privately financed investment more generally, while delivering efficient service at sustainable but just and reasonable prices.
(iii) As a consequence of the reforms, retail prices have become more closely aligned with underlying costs, and cross-subsidies have been reduced and in some countries eliminated.
(iv) There is a logical sequence of reforms, and it is costly to undertake reforms in the wrong order; ideally, the reforming country should first raise prices to cost-recovering levels (with a return on capital to finance investment), then create regulatory institutions and restructure the sector, and only after that privatize.
(v) Most developing economies lack some of the institutional and other pre-conditions for the full and effective implementation of the standard reform model. In many parts of the world, electricity markets have evolved or are evolving into hybrid forms—not completely unbundled, privatized, or competitive.
(vi) Restructuring has its own set of problems that have to be considered and weighed against the benefits of the consumers. There is not one single winning market or "one size fits all" model. In each situation, important trade-offs must be made.

X. Conclusion

This paper has outlined the reasons for power restructuring. In addition, various options which chart the evolution of the Electricity Supply Industry (ESI) from a regulated monopoly to full competition have been discussed. Four typical and successful restructured electricity markets have been explained in detail. Moreover, the status of reforms in Sub-Saharan Africa, South Asia and Latin America has been explored. Existing literature suggest that; the standard reform model has brought about a fundamental paradigm shift in terms of the structure of the ESI, the role of the state and the regulation of the sector. The generation segment of the ESI has undergone privatization in many developing countries while the network segments remain publicly owned. Overall, many developing countries are still some distance away from the full adoption of liberalized standard model in their power sector and are in transition from state control to markets.

References

[1] I. N. Kessides, "Reforming Infrastructure: Privatization, Regulation and Competition," World Bank Group and Oxford University Press, 2004.

[2] J. Kapika and A. Eberhard, Power Sector Reform and Regulation in Africa; Lessons from Kenya, Tanzania, Uganda, Zambia, Namibia and Ghana, HSRC Press, 2013.

[3] L.L. Lai, Power System Restucturing and Deregulation: Trading, Performance and Information Technology, John Wiley & Sons Ltd., 2001.

[4] J. Zhao, K. L. Lo and J. Lu, "Variously Worldwide Types of Deregulated Electricity Markets and Their Respective Transmission Congestion Management Schemes," in *UPEC 2016 - 51st International Universities' Power Engineering Conference*,, 2016.

[5] I. N. Kessides, "The Impacts of Electricity Sector Reforms in Developing Countries," *The Electricity Journal*, vol. 25, no. 6, pp. 79-88, 2012.

[6] X. Wang, "Market-Based Transmission Congestion Management Using Extended Optimal Power Flow Techniques," West London, UK, 2001.

[7] D. Kirschen and G. Strbac, Fundamentals of Power System Economics, John Wiley & Sons Ltd, 2004.

[8] J. Tooraj, N. Rabindra and G. R. Timilsina, "A quarter century yet to come of age: a survey of power sector reforms in developing countries," World Bank Group, Washington, D.C., 2015.

[9] K. N. Gratwick and A. Eberhard, "Demise of the standard model for power sector reform and the emergence of hybrid power markets," *Energy Policy*, vol. 36, pp. pp. 3948-3960, 2008.

[10] G. Simmonds, "Regulation of the UK Electricity Industry," Jan Marchant, 2002.

[11] M. Abdullah, "Congestion Management and Security Cost Allocation in Electricity Market," Ph.D. dissertation, Dept. EEE. Eng., Univ. Strathclyde, 2008.

[12] J. Sun, "Demand Side Approaches for Congestion Management in Electricity Market," Ph.D. dissertation, Dept. EEE. Eng., Univ. Strathclyde, 2014.

[13] K.L Lo and Y.S Yuen, "Deregulation of Electric Utilities," in *Power System Restructuring And Deregulation*, John Wiley Publishers, pp. 50-74.

[14] R. Dettmer, "Living with NETA [New Electricity Trading Arrangements]," *IEE Review*, vol. 48, pp. 32-36, 2002.

[15] S. Hesmondhalgh, "Is NETA the blueprint for wholesale electricity trading arrangements of the future?," *IEEE Transactions on*, vol. 18, pp. 548-554, 2003.

[16] O. Gjerde, "The deregulated Nordic electricity market-10 years of experience," in *Transmission and Distribution Conference and Exhibition 2002: Asia Pacific IEEE/PES*, 2002.

[17] L. Bergman, "The Nordic electricity market-continued success or emerging problems?," *Swedish economic policy review*, vol. 9, pp. 51-88, 2002.

[18] E. S. Amundsen and L. Bergman, "Why has the Nordic electricity market worked so well?," *Utilities Policy*, vol. 14, pp. 148-157, 2006.

[19] J. Tong, "Overview of PJM energy market design, operation and experience," in *Proceedings of the 2004 IEEE International Conference on Electric Utility Deregulation, Restructuring and Power Technologies*, 2004.

[20] A. L. Ott, "Experience with PJM market operation, system design, and implementation," *IEEE Transactions on Power Systems*, vol. 18, pp. 528-534, 2003.

[21] X. Wang, "Market-Based Transmission Congestion Management Using Extended Optimal Power Flow Techniques," PhD thesis, Department of Electronic and Computer Engineering, BruneI University, West London, UK, 2001.

[22] K. T. Madrewar, W. A. Gavhane, A. H. Kardile and U. D. Shiurkar, "Adaptive approach in deregulation of Indian power system," in *International Conference on Energy Systems and Applications*, Pune, 2015.

[23] T. Bhattacharjee and A. K. Chakraborty, "Congestion management in a deregulated power system using NSGAII," in *IEEE Fifth Power India Conference*, Murthal, 2012.

[24] K. A. Folly, *Electricity deregulation: where are we?*, ESI AFRICA, 2017.

Parametric Analysis of Fixed Bed Gasifier for Biomass and Urban Solid Wastes: A Review

Oyugi G. Oyugi, Hiram M. Ndiritu and Gathitu B. Benson

Abstract—Gasification is the most appropriate technology for converting biomass and other urban wastes into a cleaner gaseous fuels, called producer gas. Producer gas is a mixture of gases consisting of hydrogen, carbon monoxide, methane, carbon dioxide, water vapor, nitrogen, tar and suspended particulate matter. Various parametric aspects of biomass gasification have been discussed in this paper. A brief description of some of the widely used configurations of biomass gasifiers has been done. The effect of various gasifying agents on the quality of producer gas and performance of different types of fixed-bed gasifiers are discussed. The performance of the different fixed-bed biomass gasifier systems is analyzed in terms of reactor temperature, gas calorific value, cold gas efficiency, equivalence ratio, and producer gas composition. Fixed-bed gasifier has been found to produce good quality producer gas for use in chemical industries, power generation or for domestic applications

Keywords—Biomass, Fixed bed Gasifier, Performance, Producer gas.

I. INTRODUCTION

BIOMASS is a non-fossilized and biodegradable organic materials originating from plants, animals and micro-organisms. These also include products, by-products, residues and waste from agriculture, forestry and related industries, and organic industrial and municipal wastes. Biomass has been one of the main energy sources for mankind since the dawn of civilization, although its importance decreased in the late 19th century, due to increased use of oil and coal . However, recently there has been a increased interest in biomass energy in worldwide due to its benefits that include being renewable, widely available, and carbon-neutral [1].

Increased urbanization in Kenya has also resulted to the rise in the amount of solid and liquid municipals wastes generated. Many a times, landfills and dump sites have been used to manage these urban wastes. These landfills and dump sites as well as the growing industrialization of the Kenyan economy, however, have contributed immensely to the rise in the quantity of methane and heavy metals in the atmosphere and in the environment [2]. Weak implementation of laws and policies guiding waste management, and poor practices have led to towns and cities being overwhelmed by their own waste [3], thereby affecting public health and the environment.

Gasification is one of the ways through which these solid wastes and biomass, in general, can be used to generate clean energy. Gasification is a thermo-chemical process taking place at high temperatures typically above 700^oC and converts carbonaceous materials including biomass, fossil fuels, plastics, and coal into producer gas; which is a mixture of H_2, CH_4, CO, and CO_2. Limited amounts of oxygen, air, steam or their mixtures in predetermined proportions is used as the gasifying agent and heat carrier agent. The gasification of biomass into useful fuel enhances its potential as a renewable energy resource and it has emerged as a promising technology to fulfill the increasing energy demands of the world as well as to reduce significantly the volume of biomass waste generated in developing societies [4].

Oyugi G. Oyugi, Department of Mechanical and Automotive Engineering, Technical University of Mombasa (Corresponding author Phone: +254725753700; e-mail: georoys@yahoo.co.uk)).

Hiram M. Ndiritu, Department of Mechanical Engineering, JKUAT (e-mail: hndiritu@eng. jkuat.ac.ke)).

Benson B. Gathitu, Institute of Energy and Environmental Technology, JKUAT (e-mail: bbgathitu@gmail.com))).

II. GASIFICATION TECHNOLOGY OVERVIEW

A. The Stages in Gasification Process

Given that the gasifying agents and the reaction conditions are always different in the reactor during gasification, the process of biomass gasification is quite complex. The quality and efficiency of the gasification process, equipment and the products depend on the quality of the feedstock used, the composition, flow rate and the temperature of the oxidizing medium, and the prevailing temperature and pressure in the gasifier [5].

Typically gasification process consists of four stages: drying, pyrolysis, oxidation,and reduction [1], [6]–[11]:

1) *Drying*: The moisture content of biomass is typically between 5-35%. Drying occurs at about $100 - 200^oC$ with reduction in biomass moisture content falling to below 5%.

2) *Pyrolysis*: Pyrolysis is the breaking down of a biomass by heat in the absence of oxygen or air. When biomass is heated in the absence of air to about 350^oC, it forms char, hydrocarbon gases and tar vapors. The tar vapors are gases at the temperature of pyrolysis but condense to form a smoke composed of fine tar droplets as they cool. These hydrocarbon gases can condense at a sufficiently low temperature to generate liquid tars. Factors influencing devolatization include: temperature, pressure, heating rate, reaction atmosphere,and particle size of biomass.
Biomass + heat $\rightarrow$ solid, liquid, gases products (H_2, H_2O, CO, CO_2)

3) *Oxidation*: This is reaction between solid carbonized biomass and oxygen in the air, resulting in formation of CO_2 . Hydrogen present in the biomass is also oxidized to generate water. Large amount of heat is released with the oxidation of carbon and hydrogen. If the oxygen available

is in sub-stoichiometric quantities, then partial oxidation of carbon may occur, resulting in the generation of carbon monoxide hence gasification stage. This is shown in equations (1) and (2).

$$C + O_2 \overset{ExcessOxygen}{\longrightarrow} CO_2, \tag{1}$$

$$C + O_2 \overset{LessOxygen}{\longrightarrow} CO_2 \tag{2}$$

4) *Reduction*: In the absence or sub-stoichiometric presence of oxygen, several reduction reactions take place in the temperature range of between $800 - 1000^oC$. These reactions are mostly endothermic. Equations (3)-(6) show the major reactions in this stage:

$$C + H_2O \longrightarrow CO + H_2O \tag{3}$$

$$C + CO_2 \longleftrightarrow 2CO \tag{4}$$

$$CO_2 + H_2 \longleftrightarrow CO + H_2O \tag{5}$$

$$C + 2H_2 \longleftrightarrow CH_4 \tag{6}$$

B. Gasification Process Classifications and Technologies

Different methods are used in clasifying gasification processes like using the type of oxidizing agent used, using the source of heat, using gasifier operating pressure, among other methods. But the most common classification method is the use of the gasifier bed design. Under this method, gasifiers are clasified as fluidized bed, entrained bed, and fixed bed (moving bed) gasifiers. Below is a brief description of each these three types:

1) *Fluidized Bed Gasifiers* Fluidization is a process where a bed of loosely packed solid particles takes on some of the properties of a fluid when a gas is blown vertically upwards through it. A fluidized bed gasifier is a back-mixed or well-stirred reactor in which there is a consistent mixture of new and older, partially and fully gasified particles. The mixing also enhances uniform temperatures throughout the bed. The flow of gas into the reactor (oxidant, steam, recycled syngas) must be sufficient to float the particles within the bed but not so high as to entrain them out of the bed. However, as the particles are gasified, they become smaller and lighter and are entrained out of the reactor [12]. The temperatures within the bed should be less than the initial ash-fusion temperature of the particle to avoid particle agglomeration. These gasifiers are characterized by short residence time, high temperatures, high pressures, and large capacities.

2) *Entrained Flow Gasifiers*: A finely ground particle is injected in concurrent flow with the oxidant. The particle rapidly heats up and reacts with the oxidant. The residence time of an entrained flow gasifier is few to several seconds. Because of the short residence time, entrained flow gasifiers must operate at high temperatures to achieve high carbon conversion [12]. Consequently, most entrained flow gasifiers use oxygen rather than air and operate above the slagging temperature of the particle.

3) *Fixed Bed (Moving Bed) Gasifiers*: These are counter-current flow reactors in which the biomass enter from the top of the reactor and air or oxygen or steam enters at the bottom. As the particle slowly moves down through the reactor, it is gasified and the remaining ash drops out of the bottom of the reactor. Because of the counter-current flow arrangement, the heat of reaction from the gasification reactions serves to pre-heat the particle before it enters the gasification reaction zone [13]. Consequently, the temperature of the syngas exiting the gasifier is significantly lower than the temperature needed for complete conversion of the particle. Fixed bed gasifiers are simple to construct and generally operate with high carbon conversion, long feedstock residence time, low gas velocity, and low ash carry-over [14].

III. Fixed Bed Gasifiers

The choice of one type of gasifier over the other is dictated by feedstock; its final available form, size, moisture content and ash content. Fixed bed gasifiers can be classified further as updraft (countercurrent), downdraft (concurrent) or crossdraft (cross-flow). Figure 1 below shows the different types of fixed bed gasifiers.

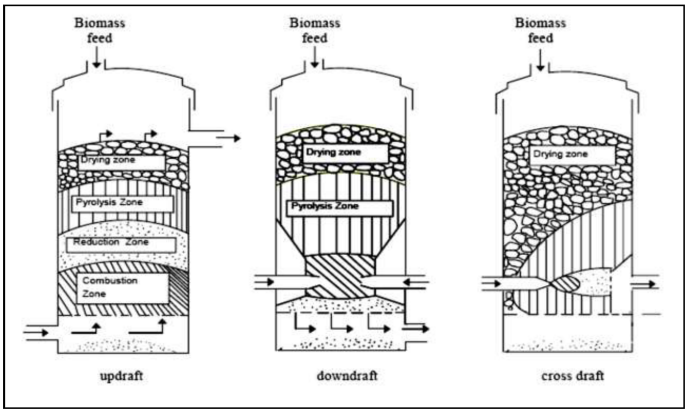


Fig. 1. Types of fixed bed gasifiers [1]

A. Updraft Gasifiers

Biomass is introduced at the top of the reactor, and a grate at the bottom of the reactor supports the reacting bed. Air or oxygen and/or steam are introduced below the grate and diffuse up through the bed of biomass and char. Complete combustion of char takes place at the bottom of the bed, liberating CO_2 and H_2O. These hot gases at about 1000^oC pass through the bed above, where they are reduced to H_2 and CO and cooled to 750^oC. Continuing up the reactor, the reducing gases (H_2 and CO) pyrolyse the descending dry biomass and finally dry the incoming wet biomass, leaving the reactor at a low temperature of about 500^oC [15].

Payne F. A., *et al* [16] studied gasification of maize cobs of varying moisture content (9-46%) in an updraft gasifier with primary air as the gasifying medium. The producer gas was then completely combusted with secondary air and the exhaust gas mixed with the ambient air and used directly for drying grain. The total particulate emission was proportional to the second power of the gasification rate (amount of corncobs gasified per unit grate area per unit time), which in turn was dependent on the corncob moisture content and primary airflow rate.

The condition of the feedstock and the rate of flow of the gasifying agents are some of the factors that affect the performance of an updraft gasifier. The producer gas from updraft gasifier has high amount of tar [17] and is therefore mostly used in thermal applications like close coupled steam boilers and crop dryers.

The main advantages of updraft gasification are:

- Simple, low cost process
- Able to handle biomass with a high moisture and high inorganic content like municipal solid waste.
- It is a proven technology for heat production
- The higher tar content in the gas adds to the heating value
- The higher operating temperatures can destroy some toxins, slag materials and some metals.
- They can be used for small-scale applications
- Exhibit very few carbon in theremaining ash.

The disadvantages of updraft gasification is:

- Syngas contains 10-20% tar by weight, requiring extensive syngas cleanup before engine, turbine or synthesis applications
- It has feedstock size limit.
- It generates a fairly low heating value gas
- The high tar content in the producer gas implies extensive producer gas cleanup before use.

B. Crossdraft gasifier

In this gasifier, biomass is introduced from the top and moves downward. The oxidizer is introduced at the bottom on one side and flows across the bed. The producer gas is extracted at $800-900^oC$ from the opposite side of the oxidizer nozzle at the same level at the grate [1]. The tar cracking capability is limited therefore the gasifier is suitable only for low-tar fuels and is mostly used in close-coupled boilers [18]. Large-scale implementation of the crossdraft gasifiers has not been done.

The advantages of cross-flow gasification include:

- It has the simplest of designs
- Exhibits stronger circulation in the hot zone of the reactor.
- The lower operating temperatures allow the use of less expensive construction materials.

The disadvantages of cross-flow gasification are:

- It is more complicated to operate.
- There are reported issues with slagging
- It produced high levels of carbon (33%) in the ash.
- The overall energy efficiency of the system is low
- The tar content in producer gas is high

C. Downdraft Gasifiers

Also known as cocurrent-flow gasification, the downdraft gasifier has the same mechanical configuration as the updraft gasifier except that the oxidant and product gases flow down the reactor, in the same direction as the biomass. A major difference is that this process can combust up to 99.9% of the tars formed. Low moisture biomass (below 20%) and air or oxygen are ignited in the reaction zone at the top of the reactor. The flame generates pyrolysis gas/vapor, which burns intensely leaving 5 - 15% char and hot combustion gas. These gases flow downward and react with the char at $800-1200^oC$, generating more CO and H_2 while being cooled to below 800^oC. Finally, unconverted char and ash pass through the bottom of the grate and are sent to disposal [15], [19].

Jayah T. H. *et al* [20] studied the gasification of chips of rubber wood of varying moisture content (12.5-18.5%) and chip size (3.35.5 cm) in an 80 kW downdraft throated gasifier. They observed that because small chips underwent faster char conversion, increasing the conversion efficiency, they therefore needed smaller gasification zone length. They further concluded that for wood chip of moisture content of 15%, the optimum chip size was 5 cm, the gasification zone length 22-33 cm, the throat angle was 610 and the carbon conversion efficiency was 56%. They observed further that the tar, water vapourand ash content were 7-9% of the producer gas output.

Kumabe K. *et al* [21] studied the co-gasification of coal and biomass with mixture of air and steam for production of syngas using a downdraft fixed bed gasifier at 900^oC. They varied the biomass to coal ratio from 0 to 1. They observed that increasing the biomass/coal ratio led to increased gas generation and decreased char and tar production. They obtained gas efficiency of 65-85%.

The main advantages of downdraft gasification are:

- Up to 99.9% of the tar formed is consumed, requiring minimal or no tar cleanup since the gas is relatively clean
- Minerals remain with the char/ash, reducing the need for a cyclone
- Proven, simple and low cost process

The disadvantages of downdraft gasification include:

- Requires feed drying to a low moisture content (less than 20%) i.e it is moisture sensitive
- Syngas exiting the reactor is at high temperature, requiring a secondary heat recovery system
- 4-7% of the carbon remains unconverted
- It generates a producer gas with low heating value
- It has limitations on feed size.

IV. Feedstocks and gasifying agents

A. Feedstocks

Different biomass feedstocks are used with fixed bed gasifiers. The major factors to consider are the biomass flow rate, type and properties. An optimum biomass flow rate is desired for the gasification system to maximize energy efficiency and it is dependent primarily on the design of the gasifier and the properties of the biomass.

The main constituents of biomass are cellulose, hemicellulose and lignin. Herbaceous crops and wood contain 60-80% cellulose and hemicellulose, and 1025% lignin [22]. The composition of these polymers in the biomass affects the product composition. Hanaoka T. *et al* [23] in their study of the effects of woody biomass components on air-steam gasification observed that at $900C$, carbon conversion efficiencies of cellulose, xylan and lignin were 97.9%, 92.2% and 52.8%, respectively. Given that fixed bed gasifiers operate at temperatures of between 800-1200^oC, they can easily be used for woody biomass gasification

Feedstocks used in gasification include wood waste (sawdust and bark), crops, agricultural waste (corn stalks), wastewater treatment plant biosolids, municipal solid waste (MSW), animal wastes (stall wastes) and blends of the various feedstocks [24]. Generally, the feedstock requires some pre-processing to remove the inorganic materials (such as metals and glass) that cannot be gasified. In addition, the MSW is typically shredded or ground into very small particles and dried before being fed into the gasifier.

Biomass may also be divided into two broad groups: (a) Virgin biomass (b) Waste. Primary or virgin biomass comes directly from plants or animals. Waste or derived biomass comes from different biomass-derived products. Table I shows the subdivisions of biomass under these two broad categories.

TABLE I
MAJOR CATEGORIES OF BIOMASS AND THEIR SUBDIVISIONS [4]

Virgin	Terrestrial biomass	Forest biomass, Grasses, Energy crops, Cultivated crops
	Aquatic biomass	Algae, Water plant
Waste	Municipal waste	Municipal solid waste Biosolids, sewage, Landfill gas
	Agricultural solid waste	Livestock and manures, Agricultural crop residue
	Forestry residues	Bark, leaves, floor residues
	Industrial wastes	Black liquor, Demolition wood Waste oil or fat

Wood: Wood fuel has several environmental advantages over fossil fuel. The main advantage is that wood is a renewable resource, offering a sustainable, dependable supply. Other advantages include the fact that the amount of carbon dioxide (CO_2) emitted during the burning process is typically 90% less than when burning fossil fuel.

Sawdust: Sawdust has a problem of excessive tar production and serious pressure drop. If the gas produced from sawdust is used in internal combustion engines, fairly good clean-up system is needed.

Peat: Peat is the first stage of coal formation. Freshly mined peat contains 90% moisture and 10% of solid. It cannot be utilized unless air-dried to reduce moisture content to 30% or less. As peat contains very high level of moisture and ash, it creates problems in the gasification process.

Agricultural residues: Agricultural residues are basically biomass materials that are by product of agriculture. It includes cotton stalks, wheat and rice straw, coconut shells, maize and rice husks and others. Coconut shells and maize cobs have been successfully tested for fixed bed gasifiers. Most cereal straws contain ash content above 10% and present slagging problem in downdraft gasifier. Rice husk with ash contents above 20% is difficult to gasify.

Municipal Solid Waste (MSW) The heat content of Refuse Derived Fuel (RDF) depends on the concentration of combustible organic materials in the waste and its moisture content. Mixed plastics and rubber contribute the highest heating values to municipal solid waste and food and yard wastes the lowest. Pre-processing MSW to RDF is an important aspect in the gasification process. If too much metal and glass are allowed to pass through into the gasifier, the heating value of the RDF decreases and there are constant operating problems. Therefore, waste gasification will be most successful in communities where there is good recycling practice, thus material recovery, by recycling, and energy recovery from waste are complementary in an integrated plan for waste management [26]

The main MSW that can be used in gasification include: a) Domestic waste, also referred to as garbage, refuse or trash consisting mainly of biodegradable waste which are food and kitchen waste, green waste paper. b) Waste tyres c) Industrial waste like waste paper, and other biodegradable industrial by-products and products, as well as plastic wastes.

B. Components of Feedstocks

The major components of any gasification feedstock are moisture, volatile matter, fixed carbon, and ash. The amount of these components vary from one feedstock to the other. examples of the composition of various feedstocks are presented in Tables II and III.

Anjireddy *et al* [4] in their study of gasification of biomass in downdraft gasifier presented a table of proximate and ultimate analysis of some biomass feedstocks as shown in Table II.

TABLE II
POTENTIAL BIOMASS GASIFIER FEEDSTOCKS [4]

	Ultimate Analysis (wt% dry basis)						Proximate Analysis (wt% dry basis)			
	C	H	N	O	S	Ash	Moisture	Volatiles	Fixed Carbon	Heating Value HHV (MJ/kg)
Agricultural Residues										
Sawdust	50	6.3	0.8	43	0.03	0.03	7.8	74	25.5	19.3
Bagasse	48	6.0	-	42	-	4	1	80	15	17
Corn Cob	49	5.4	0.4	44.6	-	1	5.8	76.5	15	17
Short Rotation Woody Crops										
Beech Wood	50.4	7.2	0.3	41	0	1.0	19	85	14	18.4
Herbaceous Energy Crops										
Switchgrass	43	5.6	0.5	46	0.1	4.5	8.4	73	13.5	15.4
Straw	43.5	4.2	0.6	40.3	0.2	10.1	7.6	68.8	13.5	17
Miscanthus	49	4.6	0.4	46	0.1	1.9	7.9	79	11.5	12
Municipal Solid Waste										
Dry Sewage	20.5	3.2	2.3	17.5	0.6	56	4.7	41.6	2.3	8
Coals										
Subbituminous	67.8	4.7	0.9	17.2	0.6	8.7	31.0	43.6	47.7	24.6
Bituminous	61.5	4.2	1.2	6.0	5.1	21.9	8.7	36.1	42.0	27.0

Raveendran *et al* [12]. obtained the composition of 13 biomass feedstocks via ultimate and proximate analysis. The result of their proximate analysis is reported in Table III.

TABLE III
PROXIMATE ANALYSIS OF BIOMASS FEEDSTOCKS [12]

Sl. no.	Biomass	VM (daf) (%)	Ash (db) (%)	Fixed carbon (100−VM) (%)
1	Bagasse	84.2	2.9	15.8
2	Coconut coir	82.8	0.9	17.2
3	Coconut shell	80.2	0.7	19.8
4	Coir pith	73.3	7.1	26.7
5	Corn cob	85.4	2.8	14.6
6	Corn stalks	80.1	6.8	19.9
7	Cotton gin waste	88.0	5.4	12.0
8	Groundnut shell	83.0	5.9	17.0
9	Millet husk	80.7	18.1	19.3
10	Rice husk	81.6	23.5	18.4
11	Rice straw	80.2	19.8	19.8
12	Subabul wood	85.6	0.9	14.4
13	Wheat straw	83.9	11.2	16.1

C. Gasifying Agents used in Fixed bed gasifiers

Optimization of the gasification operating conditions enhances production of producer gas of desired quality. It also ensures the least amount of impurities in producer gas, and increases the process net energy conversion efficiency [27].

Air: Air supplies the O_2 for combustion and effects the residence time. Controlling the air f;ow rate controls the amount of the O_2 supply which then controls the degree of combustion which in turn, affects the gasification temperature. Higher airflow rate leads in higher temperature resulting to higher biomass conversion and a higher quality of product gas. An excess degree of combustion, on the other hand, results in low energy content of the product gas as a part of biomass energy is spent during combustion. Higher airflow also shortens the residence time which may result to decreased degree of biomass conversion. [27]

Air gasification produces a poor quality gas in terms of heating values ($4 - 7MJ/m^3$ HHV) which is suitable for boiler, engine and turbine operation, but not for pipeline transportation due to its low energy density [28]. The advantages of using air include partial combustion for heat supply of gasification and moderate char and tar content in the producer gas.

Oxygen: Oxygen-to-carbon O_2/C ratio affects the operation of the gasifier very much. As the O_2/C ratio is increased, more carbon is reacted. In the same time, the ratio of CO_2 to CO increases as the O_2/C ratio approaches the stoichiometric value. This results in decrease in the heating value of the gas with increasing O_2/C ratio. The ratio of hydrogen to carbon monoxide concentration decreases with increasing O_2/C ratio [29].

Oxygen gasification produces a better quality gas (10-18 MJ/m^3 HHV) which is suitable for limited pipeline distribution and for use as synthesis gas for conversion, for example, to methanol and gasoline [28].

Steam: Supplying steam as a gasifying agent increases the partial pressure of H_2O inside the gasification chamber which favors the water gas, water gas shift and methane reforming reactions leading to increased H_2 production However, the gasification temperature needs to be high enough (above 750-800 C) for the steam reforming and water gas reactions to be favorable [30]. Catalysts can lower the operating temperature needed for the above reactions to occur.

During steam gasification, at high temperature, the H_2 yield is more pronounced than the increase in gas yield which results in an increase in H_2 content. Turn S. *et al* [31] observed that increasing Steam/Biomass from 1.1 to 4.7 decreased CO, CH_4, C_2H_2 yields, and increased H_2 and gas yields. Gupta A. K *et al* [1] observed significant increases in H_2 at temperatures above 800^oC for Steam/Biomass ratio between 0.5 and 1.08.

Since the temperature of the steam supplied to the gasifier is usually lower than the gasification temperature, and since steam has higher heat capacity, a significant amount of heat is needed to raise the steam temperature which, in turn, may lower the temperature of the gasifier bed [33]. A preheater is recommended before the introduction of gasifying agents (steam and air) to the gasifier to facilitate higher gasification bed temperature. Steam gasification produces better quality gas of almost similar HHV as oxygen gasification [28]. The advatages of using steam include production of producer gas with high heating value and rich in hydrogen gas [1].

Carbon Dioxide: In some cases carbon dioxide may be used as a gasifying agent, similar to steam. The CO_2 has an ability to convert char, tar, and CH_4 into CO and H_2, assuming the presence of catalyst, thus increasing the yield of CO and H_2 than usual [34]. The main reaction here is also endothermic, just like in case of steam, and the same means of heat input can be applied as with steam gasification. Advantages of using carbon dioxide include production of producer gas with high heating value and rich in hydrogen and carbon monoxide gases and low carbon dioxide gas [1]

NOTE: There is also a possibility of combining the above mentioned gasification agent. The most common combination is steam and air or oxygen. The agents may be introduced together in the same stage, but also separately. Indirect gasification is an example of such a process.

V. GASIFICATION PERFORMANCE PARAMETERS

The performance of gasifier is directly affected the following parameters-

A. Reactor Temperatures

Gasification temperature is one of the most influential factors affecting the product gas composition and properties and temperature is considered as the main parameter for estimating biomass gasifier performance. Increase in temperature reduces the tar content as well as decreases char inside the gasifier. It also increases gas yield due to higher tar cracking and higher conversion efficiency.

Carlos L. [35] investigated the high temperature air/steam gasification process for gasification of bark, charcoal, woodchips and wood pellets in a conventional batch type updraft gasifier. Preheated air, steam and air/steam mixture were used as gasifying media. Preheating of air up to 830^oC kept the temperature in the oxidation zone of gasifier above 1000^oC, which promoted the thermal cracking of tar in one step in

the gasifier. The lower heating value (LHV) of producer gas, cold gas efficiency and specific gas production rate (SGPR) increased from 4.6 to 7.3 MJ/Nm^3, 36.1 to 45.2% and 744.9 to 916.6 $kgm^{-2}h^{-1}$ respectively as the temperature of the preheated feed gas increased from 350 to 830^oC. With preheated air of 830^oC the LHV ($7.3MJ/Nm^3$) of producer gas was well above that reported for downdraft gasifiers. When the feed gas used was mixture of steam and air, increase in the steam fraction with respect to the high temperature air favoured the water gas shift equilibrium and steam reforming of tars and hydrocarbons which led to increase in fraction of H_2 (10.4 to 29.9%) and decrease in CO (29.4 to 18.4%) in the producer gas. However use of steam lowered the temperature of gasification and the cold gas efficiency (43 to 37%).

Gonzlez J. F. *et al* [36] observed that in air gasification, the amount of H_2 and CO in syngas increased when temperatures incresed from 700 to 900^oC, whereas contents of CH_4 and CO_2 decreased. Boateng A. A. *et al* [37] observed that increasing the gasification temperature from 700 to 800^C, increased gas yield, gas HHV, energy efficiency, carbon conversion efficiency and H2 content, while CH_4, CO and CO_2 contents decreased. The decrease in CO content may have been due to the comparatively lower temperature (than 850900 C) for the Boudouard reaction to predominate.

B. Calorific value (CV)

The calorific value (CV) of a material is an expression of the energy content, or heat value, released when burnt in air. The CV is usually measured in terms of the energy content per unit mass or volume (MJ/m^3).

Jorapur R. *et al* [38] developed a commercial scale low-density biomass gasification system, operating on sugarcane leaves, bagasse and their mixture, for thermal applications in an Agricultural Research Institute in India. The inner lining of the reactor was made of high temperature resistant firebricks. The system also had high temperature char/ash coarse settler and a high efficiency cyclone separator. They observed that the producer gas had a HHV of between 3.56-4.82 MJ/Nm^3. But the system also produced char, with gross calorific value of 18.9 MJ/kg, which was about 24% by weight of the original fuel. The performance of the gasifier was good up to fuel moisture content of 15%.

C. Cold Gas Efficiency

Cold gas efficiency is the ratio of energy of the producer gas per kg of biomass to the higher heating value (HHV) of the biomass material. Cold gas efficiency depends upon the calorific value and the amount of producer gas released at constant HHV of biomass.

$$CGE = \frac{CV \times W_{gas}}{HHV_{feed}} \tag{7}$$

where,
CGE = cold gas efficiency
CV - = Calorific value of gas,
W_{gas} - Gas production per weight of biomass.
HHV_{feed} - Higher heatingvalue of biomass.

Na I. J. *et al* [39] gasified combustible waste pellets in an updraft fixed bed gasifier lined with alumina refractory using O_2 as the gasifying medium. The composition of H_2 and CO in the producer gas was in the range 30-40% and 15-30% respectively depending on oxygen/waste ratio. Inceasing bed height increased the H_2 and CO content in producer gas while CO_2 content reduced. Gasification was difficult above a bed height of 700 mm due to pressure drop of O_2. The cold gas efficiency was around 61% at O_2/waste ratio of 0.4-0.5 and the heating value of was 11.72-13.40 MJ/Nm^3, which was more than heating value of producer gas generated by air gasification.

Dasappa *et al* [40] studied an open top ceramic-lined throatless downdraft reactor for biomass of varying moisture contents up to 37%. At a feed rate of 75 kg/h the CGE was around 75%, whereas the large capacity gasifier system of 650 kg/h resulted in cold CGE in the range of 85%. The fraction of tar and particulate matter in the hot gas from open top reactor were reported to be lower than in the throated design.

D. Equivalence ratio (λ or ER)

ER is the ratio of air flow to the airflow required for stoichiometric combustion of the biomass, which indicates extent of partial combustion. Equivalence ratio is determined from the amount of oxygen fed into the gasifier divided by the amount of oxygen required for complete fuel combustion. For the effective gasification, ER should range between 0.2 - 0.4.

$$\lambda = \frac{V_{actal}}{V_{stoic}} \tag{8}$$

where,
λ = equivalence ratio,
V_{actal} = air volume consumption by gasification,
V_{stoic} - air volume needed for complete combustion.

Sheth P.N. *et al* [41] studied a downdraft biomass gasifier used with the waste generated from sesame (rose) wood. They observed that an increase in the moisture content of the feedstock led to decrease in biomass consumption rate, while increase in the air flow rate increased the biomass consumption rate. They further observed that molar fraction of N_2 and CO_2 decreased with increase in equivalence ratio (λ) till $\lambda = 0.205$. The fraction of CO and H_2 showed increasing and decreasing trend exactly opposite to that of N_2 and CO_2. The calorific value, pyrolysis zone temperature and the oxidation zone temperature were maximum at $\lambda = 0.205$. Further they noted that the production rate of producer gas continuously increased with an increase in λ. The value of cold gas efficiency was 0.2 for $\lambda = 0.17$ and it almost doubled with a small increase of 0.035 in the value of λ.

Narvez I. *et al* [42] in their study of biomass gasification, reported that when λ was increased from 0.20 to 0.45, gas yield increased, lower heating value (LHV) of the gas decreased and the contents of H_2, CO, CH_4, C_2H_2 and tar decreased. Su Yi [43] in his study observed that, at feedstock

feed rate of around 100 kg/h and the char bed height of about 100 cm high, when λ was between 0.3 to 0.35, the heat value of producer gas reached as high as 7247.7 kJ/m^3, gas yield rate was 1.84 m^3/kg, carbon conversion rate was 91.3%, and the overall gasification efficiency was 84.6%. Wang Y. *et al* [44] in their study of woody biomass gasification found that with an increase in λ from 0.16 to 0.26, the bed and free-board temperatures rose leading to a higher yield and higher heating value (HHV) of the gas, an increase in cold gas efficiency from 57% to 74%, an increase in H_2 content from 8.5% to 13.9%, and an increase in CO content from 12.3% to 14%.

E. Producer gas composition

Producer gas composition means quantities by volume of different gases in producer gas.

Khater E. M. *et al* [**?**] studied the gasification of rice hulls in downdraft gasifiers. They used rice hulls feed rate of 1.3-5.1 kg/h and the air flow rates of 2-4.4 m^3/h. They observed that maximum yield of combustible constituents of the producer gas (13.6% CO, 5.1% H_2, 2.4% CH_4) was obtained at a stoichiometric air/fuel ratio of 55%.

Hina Beohar et al [1] conducted a study on biomass gasification on fixed bed gasifiers and presented a typical composition of producer gas obtained from downdraft gasifier with various feedstocks as shown in Table III.

TABLE IV
COMPOSITION OF SYNGAS FROM VARIOUS BIOMASS FEEDSTOCKS [1]

Fuel	Gasification method	Volume Percentage					Calorific value (MJ/m^3)
		CO	H_2	CH_4	CO_2	N_2	
Charcoal	Downdraft	28-31	5-10	1-2	1-2	55-60	4.60-5.65
Wood	Downdraft	17-22	16-20	2-3	10-15	55-50	5.00-5.86
Wheat straw pellets	Downdraft	14-17	17-19	-	11-14	-	4.50
Coconut husks	Downdraft	16-20	17-19.5	-	10-15	-	5.80
Coconut shells	Downdraft	19-24	10-15	-	11-15	-	7.20
Corn cobs	Downdraft	18.6	16.5	6.4	-	-	6.29
Rice hulls pelleted	Downdraft	16.1	9.6	0.95	-	-	3.25
Cotton stalks cubed	Downdraft	15.7	11.7	3.4	-	-	4.32

VI. CONCLUSION

There has been significance advancements in gasification processes and technologies. With increased generation of wastes in towns, need for cleaner environment, and demand for more energy, gasification technology presents an opportunity of generating clean energy from biomass and urban wastes. It helps in conserving the environment as well as mitigate supply of energy to mankind. It is obvious from the above discussion that parametric study of gasification of biomass into gaseous products using different feedstocks is feasible. The performance of the fixed bed gasifier has been discussed in terms of different parameters which significantly influence the performance of product gas quality and quantity. Increase in temperature reduces the tar content in the product gas as well as decreases char inside the gasifier. Gas yield increases due to higher tar cracking. Increase in equivalence ratio increases the temperature inside the gasifier while its decrease increases char formation inside the gasifier. From the discussion above, fixed bed gasifiers have been shown to be a simpler way to utilize biomass and municipal solid waste in generating energy.

ACKNOWLEDGMENT

We acknowledge the staff and managements of Jomo Kenyatta University of Agriculture and Technology and the Technical University of Mombasa for their support by allowing the authors to use their research facilities and resources to accomplish this review. We also thank our families for the support they gave us during this research.

REFERENCES

[1] Hina Beohar, Bhupendra Gupta*, V. K. Sethi, Mukesh Pandey, *Parametric Study of Fixed Bed Biomass Gasifier: A Review*, International Journal of Thermal Technologies, Vol.2, No.1, pp 134 140, (March 2012).

[2] Tsuma J. P. E., Wandiga S. O., Abongo D. A., *Methane and heavy metals Levels from leachates at Dandora dumpsite, Nairobi County, Kenya*, IOSR Journal of Applied Chemistry (IOSR-JAC), Vol.9, Issue 9 Ver II, pp 39 46, (September 2016).

[3] *The National Solid Waste Management Strategy*," February 2015, by national Environment management Authority, Kenya

[4] Anjireddy Bhavanam and R. C. Sastry. Biomass Gasification Processes in Downdraft Fixed Bed Reactors: International Journal of Chemical Engineering and Applications, Vol. 2, No. 6,pp 425 - 433, (December 2011).

[5] Alina Zogala, Tomasz Janoszek. *CFD simulations of influence of steam in gasification agent on parameters of UCG process*, Journal of Sustainable Mining, vol 14, pp 2-11, (2015).

[6] Beata Urych. *Determination of kinetic parameters of coal pyrolysis to simulate the process of underground coal gasification (UCG)*, Journal of Sustainable Mining vol 13(1), pp 3- 9, (2014).

[7] McKendry, P. *Energy Production from Biomass (Part 3)*: Gasification Technologies. Bioresource Technologies, vol 83(1), pp 55- 63, (2002).

[8] Li, X. *Biomass Gasification in Circulating Fluidized Bed*. Doctoral dissertation, University of British Columbia. (2002)

[9] Mandl C., Obernberger I., and Biedermann F. *Modeling of an updraft fixed-bed gasifier operated with softwood pellets*. Fuel, vol 89, pp 3795-3806, (2010).

[10] Rukayya Ibrahim Muazu, Aiduan Li Borrion, Julia A. Stegemann. *Fluidised Bed Gasification of Multiple Agricultural Biomass Derived Briquettes*, International Journal of Chemical, Molecular, Nuclear, Materials and Metallurgical Engineering vol:9 No. 5, pp 622 - 628, (2015).

[11] Knoef H. A. M. . *Handbook of Biomass Gasification,* Meppel, The Netherlands: BTG Biomass Technology Group B.V. (2005)

[12] Isam Janajreh, and Idowu Adeyemi, *Effect of Process Parameters on Gasification: A Review*, Proceedings of IAJC-ISAM International Conference, pp 1 - 13, (2014)

[13] Carlos L. *High Temperature Air/Steam Gasification of Biomass in an Updraft Fixed Bed Type Gasifier.* Doctoral dissertation, Royal Institute of Technology, Stockholm (2005)

[14] Reed T. B., and Das, A. *Handbook of Biomass Downdraft Gasifier Engine Systems*. Golden, CO: Solar Energy Research Institute. (1988)

[15] Jared P. Ciferno, John J. Marano, *Benchmarking Biomass Gasification Technologies for Fuels, Chemicals and Hydrogen Production*, National Energy Technology Laboratory, U.S. Department of Energy, pp 4 - 27, (June 2002)

[16] Payne F. A., Ross I. J., Walker J. N., and R. S. Brashear. *Exhaust analysis from gasification-combustion of corncobs*. Transactions of the ASAE, pp 246-249, (1983)

[17] Ronald W. Breault *Gasification Processes Old and New: A Basic Review of the Major Technologies*, Energies, Vol. 3, pp 216-240; (2010). doi:10.3390/en3020216

[18] Sangeeta Chopra, and Anil Jain. *A Review of Fixed Bed Gasification Systems for Biomass*, Agricultural Engineering International: the CIGR Ejournal, vol. IX, No. 5, pp 1-23, (April 2007)

[19] Younes Chhiti,Mohammed Kemiha. *Thermal Conversion of Biomass, Pyrolysis and Gasification: A Review*, The International Journal of Engineering And Science (IJES) vol:2, Issue 3, pp 75 - 85, (2013)

[20] Jayah T. H., Aye L., Fuller R. J., and Stewart D. F. *Computer simulation of a downdraft wood gasifier for tea drying*, Biomass and Bioenergy Vol. 25, pp 459-469, (2003)

[21] Kumabe K., Toshiaki Hanaoka, Shinji Fujimoto, Tomoaki Minowa and Kinya Sakanishi,*co-gasification of coal and biomass with air and steam for production of synthesis gas*, Fuel, Vol. 56, pp 684-689, (2007).
[22] Gates B.C., Huber G.W., Marshall C.L., Ross P.N., Siirola J., Wang Y. *Catalysts for emerging energy applications*, MRS Bull. vol 33, pp 429435, (2008).
[23] Hanaoka Toshiaki, Inoue S., Uno S., Ogi T., Minowa Tomoaki, *Effect of woody biomass components on air-steam gasification*, Biomass Bioenergy, vol. 28, pp 6976,(2005)
[24] *Gasification:The Waste to Energy Solution*, Gasification Technologies Council (GTC) Waste Brochure, pp 1 - 12, (August 2011), www.gasification.org
[25] Alexander Klein, and Nickolas J. Themelis, *Energy Recovery from Municipal Solid Wastes by Gasification*, North American Waste to Energy Conference 11 (NAWTEC 11) Proceedings, ASME International, Tampa Florida, pp 241-252, (April 2003).
[26] Ajay Kumar, David D. Jones, and Milford A. Hanna, *Thermochemical Biomass Gasification: A Review of the Current Status of the Technology*, Energies, 2, pp 556-581, (2009). doi:10.3390/en20300556
[27] Ajmal Latif *A study of the Design of Fluidized bed reactors for Biomass Gasification* Ph.D. thesis, Department of Chemical Engineering, University of London, London (1999)
[28] Jinsong Zhou, Qing Chen, Hui Zhao, Xiaowei Cao, Qinfeng Mei, Zhongyang Luo, and Kefa Cen. *Biomass-oxygen gasification in a high-temperature entrained-flow gasifier*, Biotechnology Advances, Vol. 27(5), pp 606-611, (2009).
[29] Kumar A., Eskridge, K., Jones, D.D., Hanna M.A. *Steam-air fluidized bed gasification of distillers grains: effects of steam to biomass ratio, equivalence ratio and gasification temperature* Bioresources Technologies, Vol. 100, pp 20622068, (2009).
[30] Turn, S., Kinoshita, C., Zhang, Z., Ishimura, D., Zhou, J. *An experimental investigation of hydrogen production from biomass gasification*, International Journal on Hydrogen Energy, 23, pp 641648, (1998).
[31] Gupta A.K., Cichonski W., *Ultrahigh temperature steam gasification of biomass and solid wastes*. Environmental Engineering Sciences, Vol. 24, pp 11791189, (2007).
[32] Lucas, C., Szewczyk, D., Blasiak, W., Mochida, S. *High-temperature air and steam gasification of densified biofuels*, Biomass Bioenergy, Vol. 27, pp 563575, (2004).
[33] Maria Puig-Arnavat, Joan Carles Bruno, and Alberto Coronas. R*eview and analysis of biomass gasification models*, Renewable and Sustainable Energy Reviews, Vol. 14 No. 9, pp 28412851, (2010).
[34] Carlos, L., *High temperature air/steam gasification of biomass in an updraft fixed batch type gasifier*, Ph.D. thesis. Royal Institute of Technology, Energy Furnace and Technology, Stockholm, Sweden, (2005).
[35] Gonzalez J.F., Roman S., Bragado D., Calderon M., *Investigation on the reactions influencing biomass air and air/steam gasification for hydrogen production*, Fuel Processing Technolologies, Vol. 89, pp 764772, (2008)
[36] Boateng A.A., Walawender W.P., Fan L.T., Chee C.S., *Fluidized-bed steam gasification of rice hull*. Bioresources Technology, Vol. 40, pp 235-239, (1992).
[37] Jorapur R. and Rajvanshi A. K. *Sugarcane leaf-bagasse gasifiers for industrial heating applications*, Biomass and Bioenergy, Vol. 13 Issue 3, pp 141-146, (1997)
[38] Na J. I., Park S. J., Kim Y. K., Lee J. G. and Kim J. H. *Characteristics of oxygenblown gasification for combustible waste in a fixed-bed gasifier*, Applied Energy Vol. 75, pp 275-285, (2003).
[39] Dasappa S., Paul P. J., Mukunda H. S., Rajan N. K. S., Sridhar G. and Sridhar H. V. *Biomass gasification technology a route to meet energy needs*. Current Science, Vol. 87 No. 7, pp 908-916, (2004).
[40] Sheth, P.N., Babu B.V. *Experimental studies on producer gas generation from wood waste in a downdraft biomass gasifier*, Bioresource Technology, Vol. 100, pp 3127-3133, (2009).
[41] Narvez I., Or, A., Aznar M.P., Corella J. *Biomass gasification with air in an atmospheric bubbling fluidized bed. Effects of six operational variables on the quality of the produced raw gas*, Ind. Eng. Chem. Res., Vol. 35, pp 21102120, (1996).
[42] Su Yi, *Experimental on rice-straw gasification in a two stage gasifier*, IEEE, (2009)
[43] Wang Y., Yoshikawa K., Namioka, T., Hashimoto Y. *Performance optimization of two-staged gasification system for woody biomass* Fuel Processing Technologies, Vol, 88, pp 243250, (2007).
[44] Khater E. M. K., El-Ibiary N. N., Khatab I. A., and Hamad M. A., *Gasification of rice hulls in downdraft gasifiers*, Biomass and Bioenergy, Vol. 3 No. 5, pp 329-333, (1992)

Architectures and Algorithms for Multiple UAV Cooperative control: A Review

Nelson Muchiri, Kamau S.I and Ikua B.W

***Abstract*—**For multiple unmanned Aerial Vehicles to collaborate in execution of a mission, a design of the mission, tasks and way points is first developed. Core to the design is determination of proper architecture and practical algorithms for the whole system. The architecture helps in integration of sensing, control, communication and planning while the algorithms are necessary for information sharing, task assignment and conflict resolution. In this paper a review of the principles of cooperative architectures and algorithm is presented. It hopes to extrapolate their importance in design hence provide essential knowledge to designers with interests in multiple vehicle cooperative control.

***Keywords*—** Control Algorithms, Cooperative control architectures, Multi Vehicle Cooperative control.

I. Introduction

APPLICATIONS of Unmanned Aerial Vehicles (UAVs) in military and civilian applications are on the rise in recent years. Such applications include exploration and mapping, search and rescue, reconnaissance, surveying, detection and monitoring in dangerous scenarios [1]. Deployment of collaborating Multiple UAVs is inevitable in most of these applications mainly because a team of cooperating UAVs offers the following advantages over single UAV [2]. First, it is possible to execute multiple simultaneous observations collect information from disparate points unlike when single UAV is deployed. Then efficiency is greatly improved by use of multiple UAVs in that tasks are executed faster. Furthermore, reliability and robustness are improved since members of the UAV team can complement each other making the entire system fault tolerant [3].

Control of multiple UAVs however is a has challenges, mainly to derive desirable collective behaviors through the design of individual agent control algorithms [4]. This has triggered research leading to the development of various architectures and algorithms for information sharing and task assignment.

II. Multi-Vehicle Cooperation Architectures

Cooperation is defined as a joint collaborative behavior that is directed toward some goal in which there is a common interest or reward [1, 5]. UAV cooperation requires the integration of sensing, control, and planning in an appropriated decisional architecture. According to wenjing, there are three cooperation architectures applied to multi-UAV research [6]. Multi-agent based architecture, work-flow based architecture and control-station based cooperation architecture.

A. Multi-agent Based Cooperation Architecture

This has been widely studied in the works of [7, 8] and by Maza [2] in which they classified various multi-agent cooperation systems based on the coupling as: Physical coupling, Formations, Swarms and Intentional cooperation.

i. Physical Coupling

Here UAVs are physically connected to each other hence their motions are constrained by forces that depend on the motion of other UAVs. This approach is mainly applied in transport problem as in the works of Bernard et-al [9, 10]. As shown in Fig 1 where a leader-follower structure is adapted for coordination.

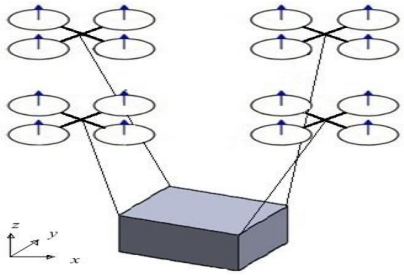


Fig 1: Load Transport with four Quadrotors

ii. Formations

Here each member of the team must keep user-defined distances with other members. Member ‘s motions are strongly constrained to keep the formation. This structure has been widely researched in the recent years and UAV aerial shapes as shown in Fig 2 have been achieved.

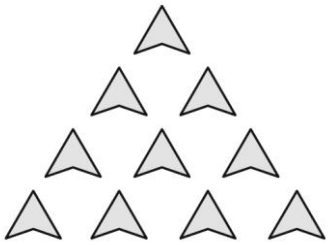


Fig 2: UAV Shape formation

For example, in research done by Turpin *et al* [12] formations strategies were developed for a team of quad-rotors following a group trajectory. The UAVs could maintain a shape or transform from one shape to another. In the works of Han *et-al* [13], Multiple UAV formations were applied for cooperative source seeking and contour mapping of a radiative signal field where they analyzed different scenarios for formation. Further research in formation has been done by Zhao *et-al* [14], where collision avoidance strategies were developed for multiple vehicles in a formation.

Nelson Muchiri, Department of Electrical Engineering, Kirinyaga University 0728073139; e-mail: ngachoki@kyu.ac.ke).

S. I. Kamau, Department of Electrical Engineering, JKUAT, Ikua B.W, Department of Mechatronic Engineering, JKUAT.

iii. Swarms

According to Maza *et al* [15], swarms are defined as teams of many vehicles in which interactions generate emerging collective behaviors. It is a concept derived from biological behaviors and typically involves a large number of homogeneous individuals [16]. Characteristics of a group of animals are imitated in coordination of a group of robots without a central supervisor, by using only local interactions between the robots as shown in Fig 3

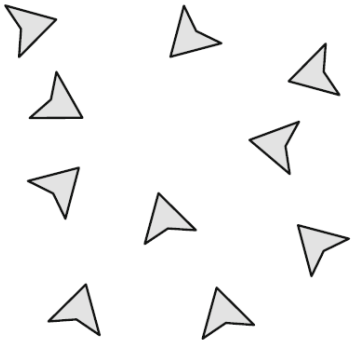


Fig 3: UAV Swarm

This has motivated development of multi-agent robotics and intelligent swarms like in the research by Vincent *et al* [17], where a framework was developed for a cooperative strategy for multiple agents searching for moving and evading targets. This was further researched by Altshuler *et al* [16] in analysis of *Cooperative Hunters Problem* where a swarm of UAVs were utilized for searching and intercepting a set of evading targets. Further applications of UAV swarms are in the works in references [18, 19, 20].

iv. Intentional Cooperation

In intentional cooperation, the UAVs of a team move in trajectories defined by individual tasks that should be allocated to perform a global mission [11]. Each UAV executes a set of tasks explicitly allocated to perform a given mission. the concept of intentional cooperation is illustrated in Fig 4. Each UAV executes its own path for accomplishment of the group objective while avoiding conflicts with other UAVs in the group.

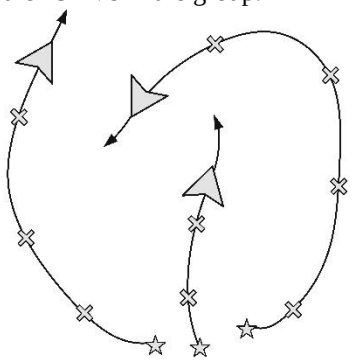


Fig 4: Intentional cooperation

This has been researched widely in works of Parker *et al* [21], Viguria *et al* [22] and Jian *et al* [8]. In design of this architecture, the main issues are task allocation, motion coordination and collision avoidance. Addressing these are emerging algorithms such as ant colony algorithm [23, 24] and others as documented in the works of Goerzen *et al* [25].

B. Other Recent Architectures

i. Hierarchical architecture

Owing to general weaknesses of the above multi-agent based cooperation architecture such as long time negotiation and reaction delays, recent research has proposed a hierarchical model of UAV coordination. developed by Pawel and Wojciech, [1] the framework is based on human organization where there is workers and superiors. The worker UAVs perform mission specific tasks such as image acquisition, while the superior UAVs are (*watching over)* supervising the function of multiple workers. This framework is illustrated in the Fig 5. Each level has its functions defined as follows, level A UAV coordinates all members of the team,

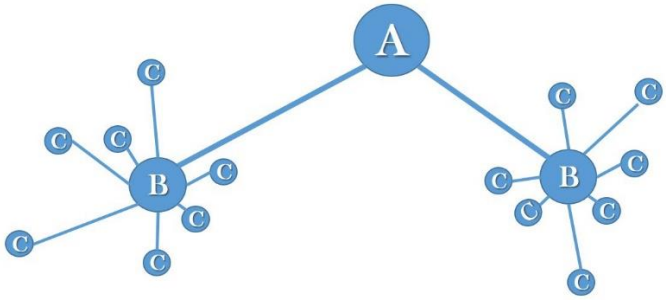


Fig 5: Hierarchical Model

level B UAV coordinates level C members in a small team and also communicates with level A. Unlike in swarms, in this model level C members cannot directly communicate with each other.

ii. Work-flow Based Architecture

This architecture has been recently proposed by Wenjin *etal* [26], where all UAVs execute tasks synchronously according to the same work flow [6]. Each UAV commands the related devices to execute tasks with the decision outputs as inputs. Therefore, because of the identical decision inputs and the same decision event, decision outputs of each UAV are same. It eliminates the need for negotiation on decision output and improves on practical time of response.

iii. Control-station Based Cooperation Architecture

With this architecture, UAVs perceive the environment, then transmit these data to the control station [6]. It has been researched previously by Jian *et al* [27] and applied in fire control by Cao *et al* [28]. The control station makes cooperative decision but UAVs make no decisions on their own.

III. Multi-vehicle Cooperation Algorithms

The principles behind operation of Cooperation algorithms is mainly dependent on the tasks being carried out by the team. For multiple air vehicles to collaborate, issues regarding information sharing, task assignment and conflict resolution are key.

A. Information Sharing

Shared information is a necessary condition for cooperation therefore information exchange among UAVs is central in their collaboration. A thorough understanding of information flow and sharing among multiple vehicles in a group is key in design of collaborative UAV systems [29]. First, the design should define the information to be communicated then decide on how to manage the information. This information may be communicated between vehicles using a wireless network, or joint knowledge might be pre-programmed into the vehicles before a mission begins [30]. The information could be about

vehicles relative position, team objectives or common control algorithms. The algorithms must make vehicles to come to an agreement commonly referred in literature as consensus. Information sharing topologies manages the exchange of local and global information among vehicles in a team [31]. The operations of team communication topology have been widely studied and can be represented by a directed graph [32, 33]. Directed graph is an approach of analyzing switching topologies by use algebraic graph theory, which associates each graph topology with an algebraic structure of corresponding matrices [29]. In these graphs, the agents are the nodes of the graph and an edge between two nodes represents an ability to communicate [34]. In a dynamic network, all nodes move and the distances between them vary with time. Such multi-agent system corresponds to a time-varying information exchange topology, and the related graph involves fixed number of nodes. In a decentralized network of multiple agents, the group performance and task accomplishment depend on agents information sharing ability to their neighbors. The algebraic connectivity, is a measure of connectivity and plays vital role in group dynamics as it determines how well the agents can communicate to each other. Team communication topologies for small teams can be demonstrated by considering a team of three members shown in Fig 6. The team members can communicate either as in (a) with a weak connection or (b) where member 1 and 2 have weak connection but member 2 and 3 have strong connection. All of them can be strongly connected as shown in (c).

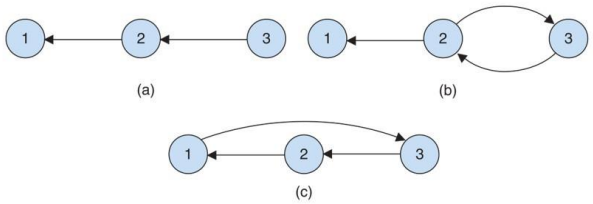


Fig 6: Different communication topologies for three vehicles

Information sharing is applied in cases of pattern formation where local information is shared between neighbors to maintain prescribed distances. In task assignment, global information helps in confirmation of overall mission.

B. Role Sharing

Another key issue in multiple vehicle collaboration is sharing of roles. Each member of a team has a role in the entire mission. The role needs to be assigned in real time, in case of a faulty member this information is shared and other members take up the role of the faulty member hence most cooperative systems are fault tolerant. If a fault happens in formation application, an algorithm repositions its neighbors to maintain the pattern. In applications where cooperative task assignment is needed, an algorithm can reassign the task aborted by a faulty member to another member in the team.

Several algorithms have been proposed for this, are broadly classified as either centralized algorithms or distributed algorithms [35]. some of the centralized algorithms include branch-and-bound procedure, enumeration method and dynamic programming. In the works of Chandler, centralized control mechanism was used to solve task allocation problem in optimization model and proposed solution approaches based on utilizing mixed integer linear programming, dynamic programming and genetic algorithms [36, 37, 38, 39]. Recent research has seen development of intelligent algorithms, such as genetic algorithm (GA) [40], particle swarm optimization (PSO), ant colony optimization (ACO) [35], wolf pack algorithm (WPA) and cat swarm optimization. Compared to distributed algorithms, the performance of the results obtained by centralized methods are generally poor in the dynamic environment.

Distributed algorithms include decentralized Markov decision process (Dec-MDP), distributed model predictive control (DMPC), dynamic distributed constraint optimization (DDCOP), contract net (CN), auction algorithm [35]. They are preferred because of their simple calculation, quick response to dynamic events, and little computation overhead and good robustness and are widely used to solve complex task allocation problems [11].

C. Consensus and Conflict Resolution

The consensus problem is to have a group of UAV reach a common assessment or decision based on distributed information and a communications protocols [31]. A consensus algorithm or protocol is an interaction rule that specifies the information exchange between an agent and all of its neighbors on the network [41]. It has been formulated as a coordinated control problem by Fax and Murray [42] where they considered a control law in which each system attempts to stabilize itself relative to its neighbors.

IV. CONCLUSION

In this paper a review of core issues in design of multiple cooperating UAV are presented. They include the principles of cooperative architectures and algorithms. The architectures are mainly multi-agent based which include formations, physical coupling and intentional cooperation. Algorithms development issues are information sharing, task assignment and conflict resolution. It is anticipated that the work presented here will elaborate critical issues in multiple UAV cooperation design hence provide essential knowledge to designers with interests in multiple vehicle cooperative control.

REFERENCES

[1] P. Rotter and W. Chmiel, “A framework for a hierarchical model of cooperation between unmanned airplanes,” *European Journal of Remote Sensing*, vol. 51, no. 1, pp. 276–284, 2018.

[2] I. Maza, A. Ollero, E. Casado, and D. Scarlatti, “Classification of multi-uav architectures,” in *Hand- book of Unmanned Aerial Vehicles*. Springer, 2015, pp. 953–975.

[3] T. H. Summers, M. R. Akella, and M. J. Mears, “Coordinated standoff tracking of moving targets: Control laws and information architectures,” *Journal of Guidance, Control, and Dynamics*, vol. 32, no. 1, pp. 56–69, 2009.

[4] J. Shamma, *Cooperative control of distributed multi-agent systems*. John Wiley & Sons, 2008.

[5] T. Shima and S. Rasmussen, *UAV cooperative decision and control: challenges and practical ap- proaches*. SIAM, 2009.

[6] Wenjing and X. Shenghong, “Comparison of multi-uav cooperation architectures,” in *Information Management (ICIM), 2017 3rd International Conference on*. IEEE, 2017, pp. 500–505.

[7] L. Merino, F. Caballero, J. R. Mart´ınez-de Dios, J. Ferruz, and A. Ollero, “A cooperative perception system for multiple uavs: Application to automatic detection of forest fires,” *Journal of Field Robotics*, vol. 23, no. 3-4, pp. 165–184, 2006.

[8] J. Han, C.-h. Wang, and G.-x. Yi, “Cooperative control of uav based on multi-agent system,” in *Industrial Electronics and Applications (ICIEA), 2013 8th IEEE Conference on*. IEEE, 2013, pp. 96–101.

[9] M. Bernard, K. Kondak, I. Maza, and A. Ollero, "Autonomous transportation and deployment with aerial robots for search and rescue missions," *Journal of Field Robotics*, vol. 28, no. 6, pp. 914–931, 2011.
[10] S. Aghdam, M. B. Menhaj, F. Barazandeh, and F. Abdollahi, "Cooperative load transport with movable load center of mass using multiple quadrotor uavs," in Control, Instrumentation, and Au- tomation (ICCIA), 2016 4th International Conference on. IEEE, 2016, pp. 23–27.
[11] M.-H. Kim, H. Baik, and S. Lee, "Response threshold model based uav search planning and task allocation," *Journal of Intelligent & Robotic Systems*, vol. 75, no. 3-4, pp. 625–640, 2014.
[12] M. Turpin, N. Michael, and V. Kumar, "Decentralized formation control with variable shapes for aerial robots," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 23–30.
[13] J. Han and Y. Chen, "Multiple uav formations for cooperative source seeking and contour mapping of a radiative signal field," *Journal of Intelligent & Robotic Systems*, vol. 74, no. 1-2, pp. 323–332, 2014.
[14] I. Zhao, S. Li, J. Hu, Q. Pan, Z. Xu, Y. Lvy, and W. Meng, "A study on uav formation collision avoidance," in *Control Conference (CCC), 2017 36th Chinese*. IEEE, 2017, pp. 8743–8747.
[15] Maza, J. Capit´an, L. Merino, and A. Ollero, "Multi-uav cooperation," *Encyclopedia of Aerospace Engineering*, 2016.
[16] Y. Altshuler, V. Yanovsky, I. A. Wagner, and A. M. Bruckstein, "Efficient cooperative search of smart targets using uav swarms," *Robotica*, vol. 26, no. 4, pp. 551–557, 2008.
[17] P. Vincent and I. Rubin, "A framework and analysis for cooperative search using uav swarms," in *Proceedings of the 2004 ACM symposium on Applied computing*. ACM, 2004, pp. 79–86.
[18] J. Hu, J. Xu, and L. Xie, "Cooperative search and exploration in robotic networks," *Unmanned Systems*, vol. 1, no. 01, pp. 121–142, 2013.
[19] Undeger and F. Polat, "Multi-agent real-time pursuit," *Autonomous Agents and Multi-Agent Systems*, vol. 21, no. 1, pp. 69–107, 2010.
[20] Y. Altshuler, A. Pentland, and A. M. Bruckstein, "The cooperative hunters–efficient and scalable drones swarm for multiple targets detection," in *Swarms and Network Intelligence in Search*. Springer, 2018, pp. 187–205.
[21] L. E. Parker, "Distributed intelligence: Overview of the field and its application in multi-robot systems."
[22] A. Viguria, I. Maza, and A. Ollero, "Distributed service-based cooperation in aerial/ground robot teams applied to fire detection and extinguishing missions," *Advanced Robotics*, vol. 24, no. 1-2, pp. 1–23, 2010.
[23] W. Cao and W. Yang, "A survey of vehicle routing problem," in *MATEC Web of Conferences*, vol.100. EDP Sciences, 2017, p. 01006.
[24] M. Dorigo and L. M. Gambardella, "Ant colony system: a cooperative learning approach to the traveling salesman problem," *IEEE Transactions on evolutionary computation*, vol. 1, no. 1, pp. 53– 66, 1997.
[25] Goerzen, Z. Kong, and B. Mettler, "A survey of motion planning algorithms from the perspective of autonomous uav guidance," *Journal of Intelligent and Robotic Systems*, vol. 57, no. 1-4, p. 65, 2010.
[26] W. Cao and S. Xu, "A new multi-uav cooperation method," in *Computational Intelligence and Design (ISCID), 2016 9th International Symposium on*, vol. 2. IEEE, 2016, pp. 231–234.
[27] J. Li, X. Fu, and X. Gao, "Cooperative multi-uav path planning with communication constraints, "*Electronics Optics & Control*, vol. 20, no. 6, pp. 29–33, 2013.
[28] J. Cao and X. Gao, "Agent-based design for multi-ucav intelligent command and control cooperative system," *Fire Control & Command Control*, vol. 28, no. 5, pp. 22–24, 2003.
[29] W. Ren and R. W. Beard, *Distributed consensus in multi-vehicle cooperative control*. Springer, 2008.
[30] W. Ren, R. W. Beard, and E. M. Atkins, "A survey of consensus problems in multi-agent coordina- tion," in *American Control Conference, 2005. Proceedings of the 2005*. IEEE, 2005, pp. 1859–1864.
[31] R. M. Murray, "Recent research in cooperative control of multivehicle systems," *Journal of Dynamic Systems, Measurement, and Control*, vol. 129, no. 5, pp. 571–583, 2007.
[32] W. Ren, R. W. Beard, and E. M. Atkins, "Information consensus in multivehicle cooperative control," *IEEE Control Systems*, vol. 27, no. 2, pp. 71–82, 2007.
[33] C. Godsil and G. F. Royle, *Algebraic graph theory*. Springer Science & Business Media, 2013, vol.207.
[34] R. Dutta, "Cooperative control of autonomous network topologies," Ph.D. dissertation, The Univer- sity of Texas at San Antonio, 2016.
[35] H. Wu, H. Li, R. Xiao, and J. Liu, "Modeling and simulation of dynamic ant colony's labor division for task allocation of uav swarm," *Physica A: Statistical Mechanics and its Applications*, vol. 491, pp. 127–141, 2018.
[36] K. E. Nygard, P. R. Chandler, and M. Pachter, "Dynamic network flow optimization models for air vehicle resource allocation," in *American Control Conference, 2001. Proceedings of the 2001*, vol. 3. IEEE, 2001, pp. 1853–1858.
[37] Schumacher, P. Chandler, M. Pachter, and L. Pachter, "Uav task assignment with timing con- straints via mixed-integer linear programming," in *AIAA 3rd" Unmanned Unlimited" Technical Conference, Workshop and Exhibit*, 2004, p. 6410.
[38] T. Shima, S. J. Rasmussen, A. G. Sparks, and K. M. Passino, "Multiple task assignments for co- operating uninhabited aerial vehicles using genetic algorithms," *Computers & Operations Research*, vol. 33, no. 11, pp. 3252–3269, 2006.
[39] V. K. Shetty, M. Sudit, and R. Nagi, "Priority-based assignment and routing of a fleet of unmanned combat aerial vehicles," *Computers & Operations Research*, vol. 35, no. 6, pp. 1813–1828, 2008.
[40] M. Darrah, W. Niland, B. Stolarik, and L. Walp, "Uav cooperative task assignments for a sead mission using genetic algorithms," in *AIAA Guidance, Navigation, and Control Conference and Exhibit*, 2006, p. 6456.
[41] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, 2007.
[42] J. A. Fax and R. M. Murray, "Information flow and cooperative control of vehicle formations," *IEEE transactions on automatic control*, vol. 49, no. 9, pp. 1465–1476, 2004.

Dual-band fractal antenna with concentric ring-shaped defected ground plane structure

Edwin Kimani Miring'u*, Kibet Langat and D. B. O. Konditi

***Abstract*—** This paper proposes the design of improved dual-band micro-strip antennas through the use of concentric rectangular fractals in the patch, and a ground plane with a concentric ring-shaped defected ground plane structure, located at the bottom of the same antenna. This design technique constitutes the fractal antenna with concentric ring-shaped defected ground plane structure. The design will be undertaken in two successive iterations. The antennas proposed in this paper have dual-band resonances at 1.7375 GHz and 2.0375 GHz for the 1st iteration, and 1.7625 GHz and 2.075 GHz for the 2nd iteration, with -10 dB impedance bandwidths of 72.5 MHz and 65.5 MHz for the 1st iteration, and 58.8 MHz and 66.8 MHz for the 2nd iteration. The designed antennas can be used for telecommunication and satellite applications.

***Keywords*—** Defected, fractal, micro-strip, multi-band.

I. INTRODUCTION

Due to the proliferation of wireless communication systems, micro-strip antennas have found increased use in these systems due to their inherent advantages such as light weight, ease off fabrication, low cost and conformability to mounting surfaces [1].

Micro-strip antennas are however faced by the disadvantage of only radiating efficiently over a narrow band of frequencies, which in turn limits the band of frequencies over which they can operate satisfactorily. The rapid development of modern mobile and wireless communication systems has thus necessitated the design of micro-strip antennas with multiband capability [2].

Various design techniques have been employed by various authors to achieve multi-band operation such as defected ground plane structure (DGPS) and use of fractal geometries in patches.

A DGPS is a simply a defect or defects excoriated in the ground plane of the antenna. The DGPSs are easy to design and implement. They are however not very complex in shape to prevent difficulty during the optimization process when particular design goals are required [3]-[9].

Fractal antennas, on the other hand, are recursively generated geometries based on the concept of a fractal. They are generated through an iterative process which leads to self-similar structures. Fractals exist in nature and can be used to model complex phenomena such as mountains, trees, clouds etc [10]. Various fractal design techniques have been carried out to improve the performance of micro-strip antennas [11]-[19].

This paper proposes the design of improved dual-band micro-strip antennas through excoriating concentric rectangular fractals in the patch and a ground plane with a concentric ring-shaped defected ground plane structure at the bottom of the same antenna. The substrate has a relative permittivity $\varepsilon_r = 4.7$ and a thickness, h of 1.6 mm.

In a previous study [20], a fractal antenna was designed, which exhibited single band resonance. This paper proposes to incorporate concentric rectangular fractals in the patch, in addition to concentric ring-shaped defects in the ground plane of the same antenna, which constitutes the fractal antenna with concentric ring-shaped defected ground plane structure (FACRDGPS). The design will be undertaken in two successive iterations (1st and 2nd).

The antennas proposed in this paper exhibit dual-band resonance at 1.7375 GHz and 2.0375 GHz for the 1st iteration, and 1.7625 GHz and 2.075 GHz for the 2nd iteration, with -10 dB impedance bandwidths of 72.5 MHz and 65.5 MHz for the 1st iteration, and 58.8 MHz and 66.8 MHz for the 2nd iteration.

The designed antennas can be used for mobile communication and satellite radio applications which require multi-band antenna for their applications and since the resonance frequencies fall in the Ultra High Frequency (UHF) band (0.3-3GHz).

II. ANTENNA DESIGN

A. Conventional inset-fed rectangular patch design

In the design of the Fractal antenna with Concentric Ring-shaped Defected ground plane Structure (FACRDGPS), a conventional rectangular micro-strip antenna is initially designed. A resonance frequency of 2.45 GHz is chosen because of its location in the Industrial, Scientific and Medical (ISM) band and also due to its low attenuation through the atmosphere.

The substrate used is FR4 with a dielectric constant ε_r of 4.7, a loss tangent of 0.02 and a thickness h of 1.6 mm. FR4 is chosen for the design due to its low cost.

A micro-strip feed line is used as the feeding technique because it is simple to match, easy to fabricate and simple to model [21]. An inset feed is used to match the patch antenna's

Edwin Kimani Miring'u, Department of Telecommunication and Information Engineering, JKUAT (corresponding author to provide phone: +254 727177541; e-mail: miringu4edwin@gail.com).

author@eng. jkuat.ac.ke).

input impedance to that of the micro-strip line (50 Ω). The conventional rectangular patch is shown in Fig.1.

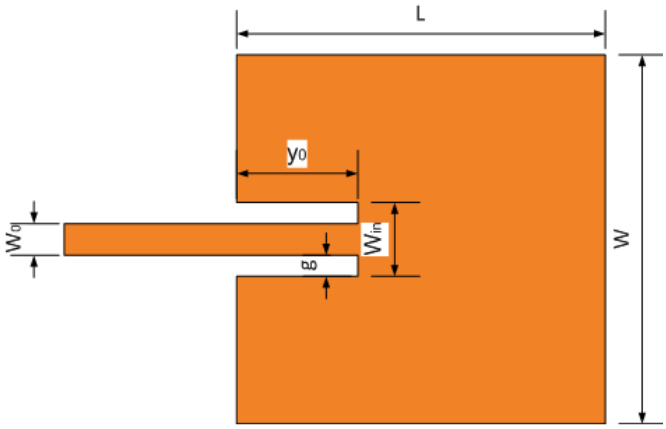


Fig. 1 An inset fed conventional rectangular patch antenna

The width W of the antenna in Fig. 1 is computed using equation (1) [1];

$$W = \frac{c}{2f_r}\left[\frac{\varepsilon_r+1}{2}\right]^{-0.5} \quad (1)$$

where c is the speed of light in m/s, f_r is the resonance frequency in Hz, and ε_r is the dielectric constant.

The effective dielectric constant, ε_{reff} is then computed using equation (2) [1];

$$\varepsilon_{reff} = \frac{\varepsilon_r+1}{2} + \frac{\varepsilon_r-1}{2}\left[1+\frac{12h}{W}\right]^{-0.5} \quad (2)$$

where h is the height/thickness of the substrate in m.

The extension of the length, ΔL due to fringing of electric fields is then computed using equation (3) [1];

$$\Delta L = 0.412h\frac{[\varepsilon_{reff}+0.3]}{[\varepsilon_{reff}-0.258]}\frac{\left[\frac{W}{h}+0.264\right]}{\left[\frac{W}{h}+0.8\right]} \quad (3)$$

The effective length of the patch L_{eff} is then computed using equation (4) [1];

$$L_{eff} = \frac{c}{2f_r\sqrt{\varepsilon_{reff}}} \quad (4)$$

The actual length of the patch L is then computed using equation (5) [1];

$$L = L_{eff} - 2\Delta L \quad (5)$$

The ground plane length L_g and width W_g are then computed using equation (6) and (7) respectively [1];

$$L_g = L + 12h \quad (6)$$

$$W_g = W + 12h \quad (7)$$

The characteristic impedance Z_c of the micro-strip line feed is given by equation (8). From equation (8), the width of the micro-strip line feed is computed [1].

$$Z_c = \frac{120\pi}{\sqrt{\varepsilon_{reff}}}\left[\frac{W_0}{h} + 1.393 + 0.667\ln\left(\frac{W_0}{h}+1.444\right)\right] \quad (8)$$

where W_0 is the width of the micro-strip line feed in m. Since equation (8), doesn't have a closed form solution, a MATLAB code was used to compute W_0.

The inset width W_{in} as shown in Fig.1, is calculated using equation (9) [22];

$$W_{in} = \frac{W}{5} \quad (9)$$

The gap g of the inset feed is then calculated using equation (10) [22];

$$g = \frac{W_{in}-W_0}{2} \quad (10)$$

The input resistance R_{in} for the inset feed is given by equation (11) [1];

$$R_{in}(y = y_0) = R_{in}(y = 0)\cos^2\frac{\pi}{L}y_0 \quad (11)$$

where y_0 is the inset feed distance in m. A MATLAB code was used to compute the input resistance at the leading radiating edge of the antenna. The same MATLAB code was also used to compute the inset feed distance y_0.

After the initial design computations, the antenna design and simulation are carried out using High Frequency Structure Simulator (HFSS) simulation software. Optimization of the patch length L and width W is done for resonance at exactly 2.45 GHz. The optimized patch length and width dimensions are denoted by L_{opt} and W_{opt}, respectively in Table I.

The computed parameters of the inset fed-rectangular patch antenna are indicated in Table I.

Table I Dimensions of the conventional inset-fed rectangular patch

Antenna Parameter	Notation	Dimension
Calculated patch	$L \times W$	27.91×36.27
Optimized patch	$L_{opt} \times W_{opt}$	28 × 34
Ground plane	$L_g \times W_g$	47.10×55.47
Substrate	$L_s \times W_s$	47.10×55.47
Inset feed distance	y_0	10.42
Micro-strip line width	W_0	3.2
Notch width	g	2.03

B. Fractal antenna with concentric ring-shaped defected ground plane structure (FACRDGPS) design

In designing the Fractal Antennas with Concentric Ring-shaped DGPS (FACRDGPS), a defected ground plane structure

(DGPS) based on concentric rings was incorporated in the ground plane, in the 1st and 2nd iterations of a Fractal antenna. Thus the 1st iteration FACRDGPS is designed and subsequently, the 2nd iteration FACRDGPS is designed.

i. 1st iteration FACRDGPS Design

The 1st iteration FACRDGPS incorporates a 1st iteration fractal in the patch of the antenna and a concentric ring shaped DGPS in the ground plane of the antenna.

The 1st iteration Fractal antenna is shown in Fig. 2. For the 1st iteration fractal, two concentric rectangles are used in generating the fractal. The main patch is divided into 9 rectangles each of dimensions $L_{01} \times W_{01}$. L_{01} is $\frac{1}{3}$ of L whereas W_{01} is $\frac{1}{3}$ of W.

Another rectangle with the dimensions $L_{i1} \times W_{i1}$ is designed. L_{i1} is $\frac{1}{4}$ of L whereas W_{i1} is $\frac{1}{4}$ of W. The subscripts 0 and i imply the outer and inner rectangle, respectively. The subscript 1 implies the 1st iteration. Thus the outer rectangle has the dimensions $L_{01} \times W_{01}$ being equal to 9.3333×11.3333 mm^2, whereas the inner rectangle has the dimensions $L_{i1} \times W_{i1}$ being equal to 7×8.5 mm^2.

Using the two concentric rectangles of dimensions $L_{01} \times W_{01}$ at the center of the patch, which is the central rectangle of the 9 rectangles that the patch was sub-divided into, Boolean subtraction is carried out using the HFSS simulation software. This gives rise to the concentric fractal rectangle as depicted in Fig. 2. This constitutes the 1st iteration fractal antenna shown in Fig. 2.

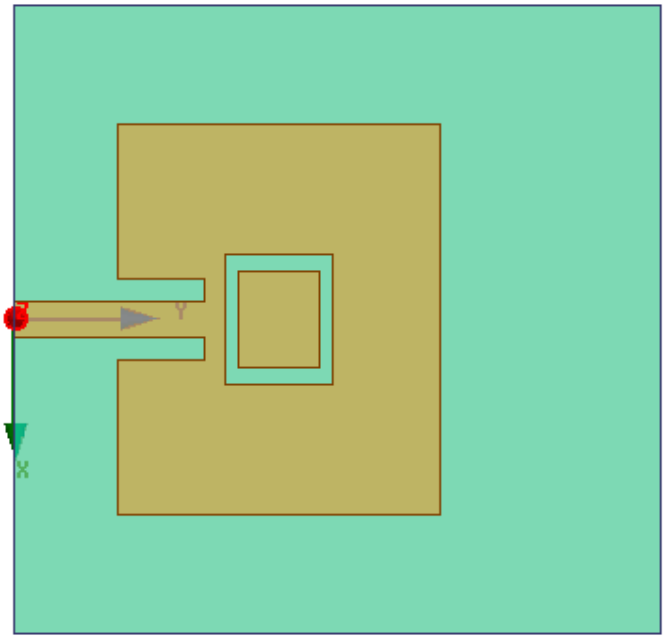


Fig. 2 1st iteration fractal antenna

The concentric ring-shaped defected ground plane structure (CRDGPS), based on concentric circles, is excoriated in the ground plane of the antenna to form the concentric ring-shaped defected ground plane structure micro-strip antenna (CRDGPSMA). The dimensions of the CRDGPS are indicated in Table II. The CRDGPSMA is depicted in Fig. 3.

Table II Dimensions of the concentric ring-shaped defected ground plane structure (CRDGPS)

Dimensions of the CRDGPS (mm)							
R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
8.5	7	11.5	10	14.5	13	17.5	16

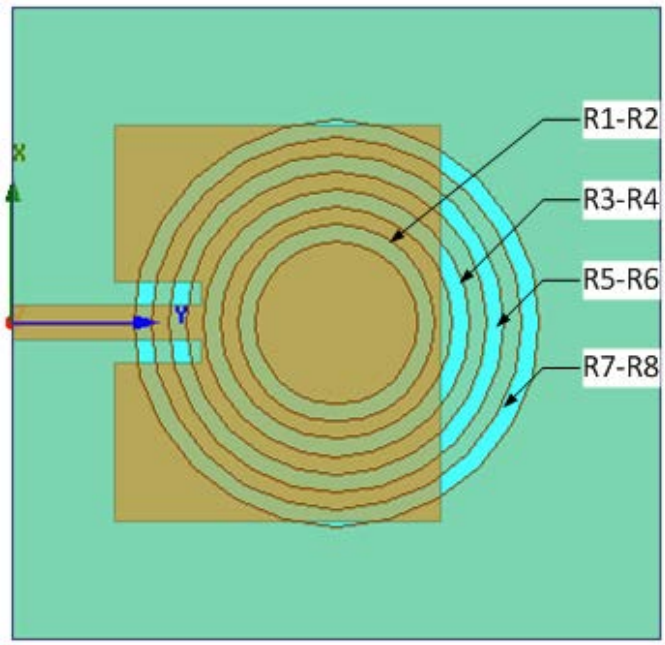


Fig. 3 The concentric ring shaped defected ground plane structure incorporated in the ground plane of the antenna

By incorporating both design techniques in the same antenna as detailed above; the 1st iteration fractal in the patch of the antenna and excoriating the concentric ring-shaped defected ground plane structure in the ground plane, a 1st iteration fractal antenna with concentric ring-shaped defected ground plane structure (1st iteration FACRDGPS) is designed. The 1st iteration FACRDGPS is depicted in Fig. 4.

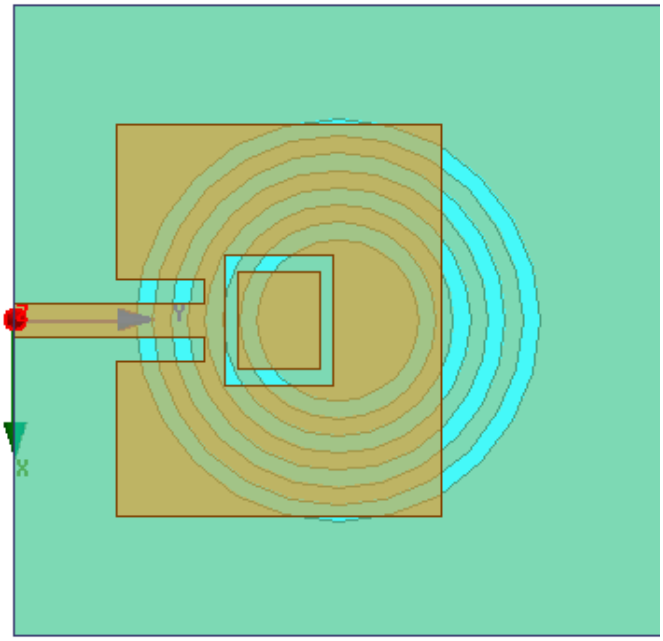


Fig. 4 1st iteration Fractal antenna with Concentric Ring-shaped DGPS (1st iteration FACRDGPS)

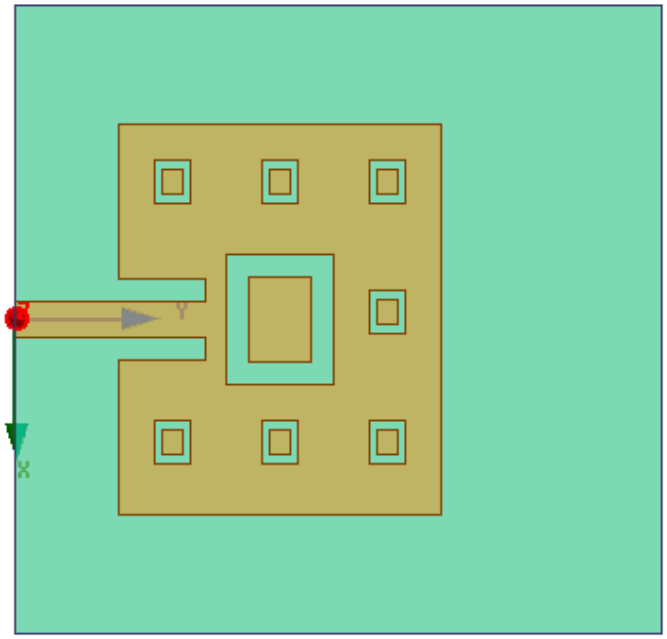


Fig. 5 The 2nd iteration fractal antenna

ii. 2nd iteration FACRDGPS

In designing the 2nd iteration Fractal antenna, the 1st iteration fractal antenna designed in the section above and shown in Fig. 2 is used.

Each of the 9 rectangles into which the main patch was initially sub-divided, is further sub-divided into 9 rectangles. The dimensions of each of these small rectangles are $L_{02} \times W_{02}$. L_{02} is $\frac{1}{3} L_{01}$ whereas W_{02} is $\frac{1}{3} W_{01}$.

An additional rectangle is created with the dimensions $L_{i2} \times W_{i2}$. L_{i2} is $\frac{1}{4} L_{i2}$ whereas W_{i2} is $\frac{1}{4} W_{i1}$. The subscripts o and i imply the outer and inner rectangle, respectively. The subscript 2 denotes the 2nd iteration.

Thus, the outer rectangle had the dimensions $L_{02} \times W_{02}$ equal to 3.1111×3.7778 mm^2, whereas the inner rectangle had the dimensions $L_{i2} \times W_{i2}$ equal to 1.75×2.125 mm^2.

Using the two concentric rectangles of dimensions $L_{02} \times W_{02}$ and $L_{i2} \times W_{i2}$ at the center of each of the 9 sub-rectangles that each of the 9 rectangles in the 1st iteration fractal is subdivided into, Boolean subtraction is carried out using the HFSS simulation software, to form 7 smaller concentric fractal rectangles. One set of concentric fractal rectangle couldn't be excoriated due to the location of the micro-strip feed line. This gives rise to the fractal concentric rectangles as depicted in Fig. 5. This constituted the 2nd iteration fractal antenna, shown in Fig. 5.

The 2nd iteration fractal antenna with concentric ring-shaped defected ground plane structure (FACRDGPS) design involves excoriating 2nd iteration fractals in the patch and concentric ring-shaped defects in the ground plane of the antenna. The concentric ring-shaped defected ground plane structure to be used is the same as the one depicted in Fig. 3. Combining these two design techniques results in a 2nd iteration FACRDGPS. The 2nd iteration FACRDGPS is illustrated in Fig. 6.

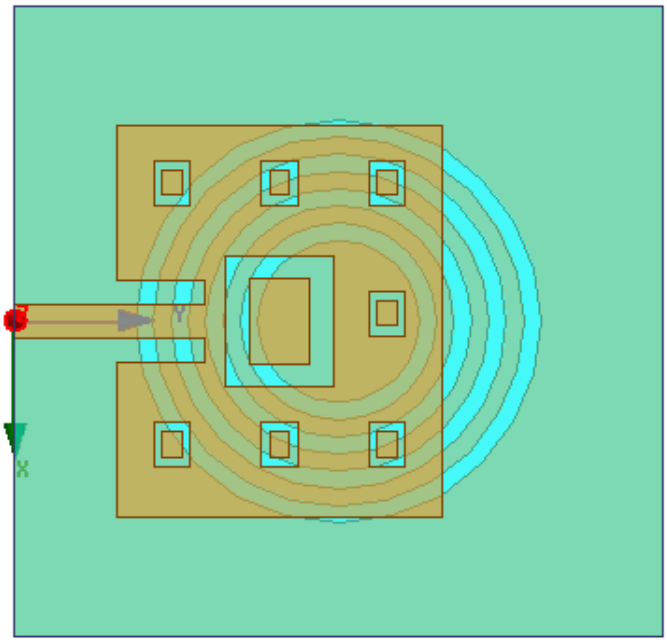


Fig. 6 2nd iteration fractal antenna with concentric ring-shaped defected ground plane structure (FACRDGPS)

III. RESULTS AND DISCUSSION

A. Conventional inset fed rectangular patch

The return loss plot of the conventional inset fed rectangular patch is shown in Fig. 7. For good power coupling, the return loss should be less than -10 dB. The return loss for the conventional inset fed rectangular patch is -21.5064 dB at a resonance frequency of 2.45 GHz with -10 dB impedance bandwidth was 67.5 MHz.

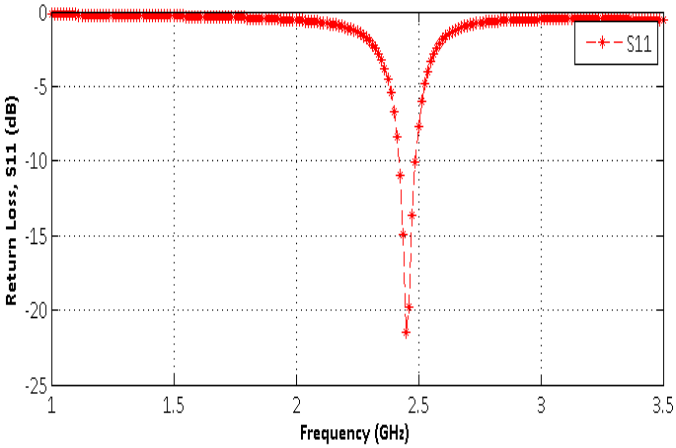


Fig. 7 Return loss plot of the conventional inset fed rectangular patch.

The Voltage Standing Wave Ratio (VSWR) plot is depicted in Fig. 8. Across the band from 2.4190 GHz--2.4896 GHz, the VSWR is observed to be < 2. This is indicative of good impedance matching which prevents power from being reflected back to the source, which would lead to power losses.

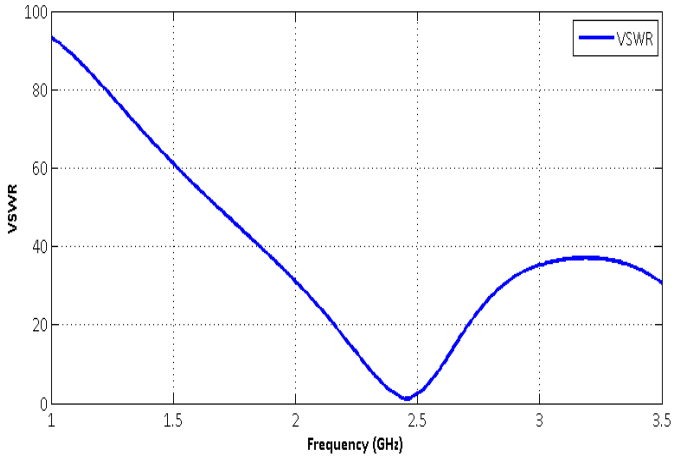


Fig. 8 VSWR plot of the conventional inset fed patch antenna

In order to study the excitation mechanism of the conventional inset fed rectangular patch, the current distribution of the antenna is studied. This is shown in Fig. 9. Fig. 9(a) shows the magnitude form of the current distribution whereas Fig. 9(b) shows the current distribution in vector form. The current distribution is observed to be concentrated along the mid-section of the antenna. This can be attributed to the fact that most of the radiation occurs from this section.

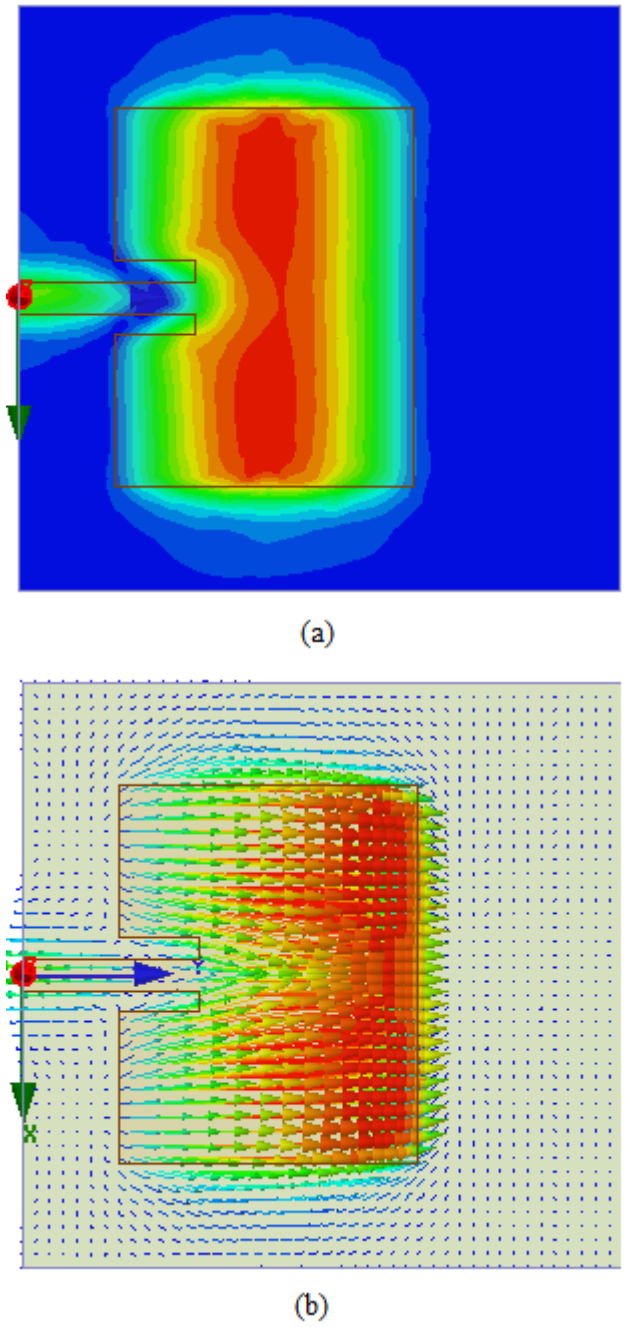


Fig. 9 Current distribution of the conventional patch antenna at 2.45 GHz; (a) Magnitude form (b) Vector form

From the current distribution plot (in vector form) of the conventional patch antenna shown in Fig. 9(b), it is observed that current flows in the *y*-direction. Thus, the *yz*-plane is the *E*-plane whereas the *xz*-plane is the *H*-plane. The radiation patterns for the conventional patch antenna in the *E*-plane and the *H*-plane at 2.45 GHz are plotted in Fig. 10.

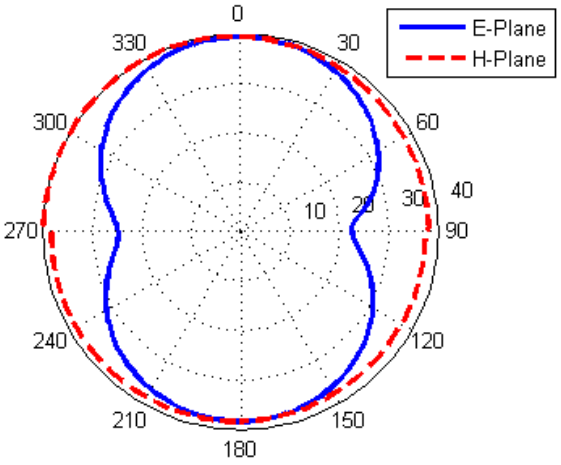


Fig. 10 Radiation pattern of the conventional patch antenna at 2.45 GHz

The radiation pattern in the *E*-plane is observed to be bi-directional whereas the one in the *H*-plane was observed to be omni-directional. The radiation patterns are observed to be symmetrical due to the symmetrical structure of the conventional patch antenna.

B. 1st iteration FACRDGPS Design

The return loss plots of the 1st iteration Fractal antenna with concentric ring-shaped DGPS (FACRDGPS) are shown in Fig. 11. The return loss plots of the Concentric Ring-shaped DGPSMA and fractal iteration 1 antenna are included for comparison.

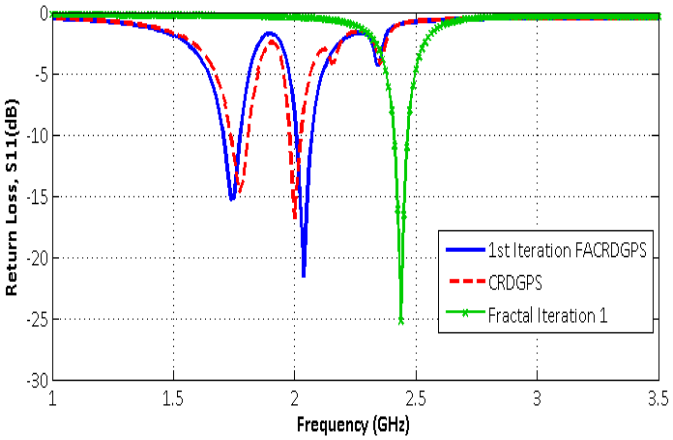
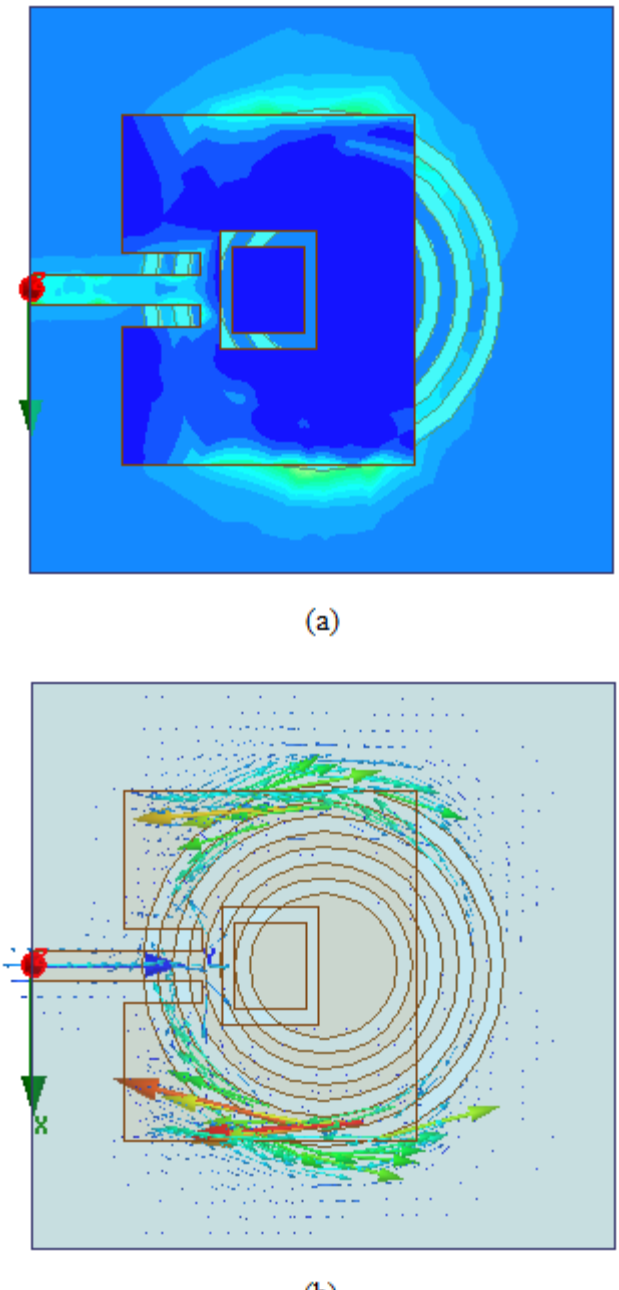


Fig. 11 Combined return loss plot of the 1st iteration FACRDGPS, Concentric Ring-shaped DGPSMA and the 1st iteration Fractal antenna

The 1st iteration FACRDGPS exhibits dual-band resonance, resonating at 1.7375 GHz and 2.0375 GHz, with –10 dB impedance bandwidths of 72.5 MHz and 65.5 MHz, respectively. The return losses (S_{11}) were -15.2374 dB and -21.6264 dB.

The current distributions of the 1st iteration FACRDGPS are depicted in Fig. 12 and Fig. 13 at resonances of 1.7375 GHz and 2.0375 GHz, respectively.



(a)

(b)

Fig. 12 Current distribution of the 1st iteration fractal antenna with concentric ring-shaped DGPS (FACRDGPS) at 1.7375 GHz; (a) Magnitude (b) Vector

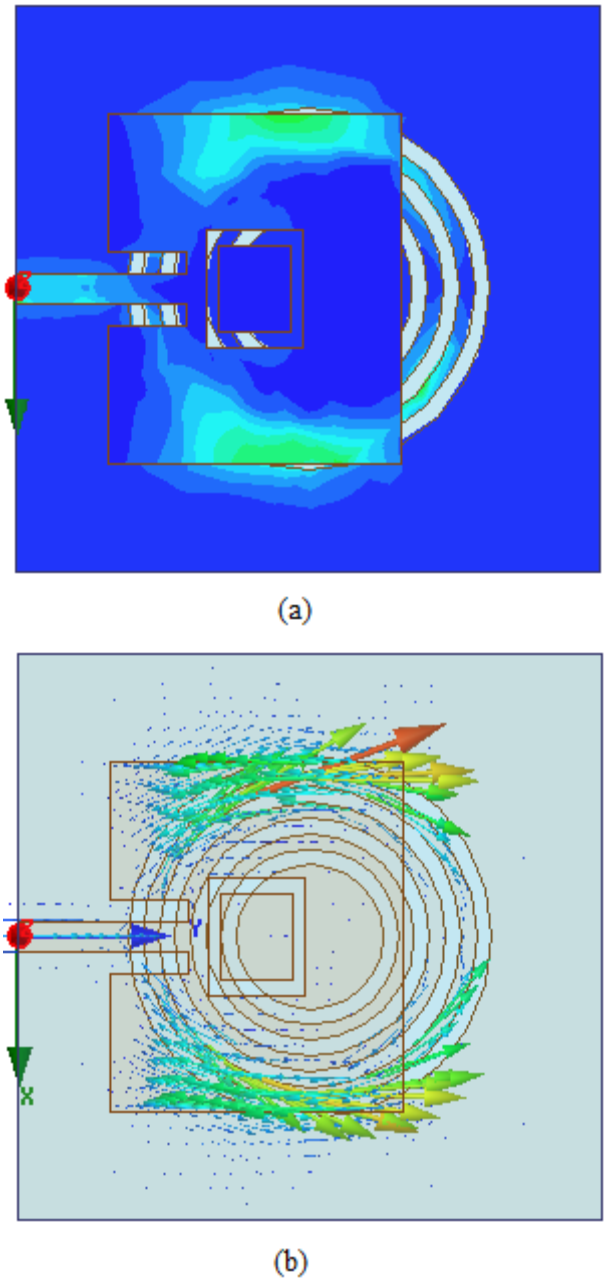


Fig. 13 Current distribution of the 1st iteration fractal antenna with concentric ring-shaped DGPS (FACRDGPS) at 2.0375 GHz; (a) Magnitude (b) Vector

From Fig. 12(b) and Fig. 13(b) for the 1st iteration FACRDGPS, compared to Fig. 9(b) of the conventional patch antenna, it can be observed that the disturbance of the current distribution caused by the excoriation of the fractals in the patch and the concentric ring-shaped defects in the ground plane causes the current to flow in two different paths. This leads to the dual-band resonance observed at 1.7375 GHz and 2.0375 GHz.

The radiation patterns of the 1st iteration FACRDGPS are plotted in Fig. 14 and Fig. 15, at 1.7375 GHz and 2.0375 GHz, respectively. From the current distribution plot (in vector form) of the 1st iteration FACRDGPS at 1.7375 GHz shown in Fig. 12(b), it is observed that current flows in the *y*-direction. Thus, the *yz*-plane is the *E*-plane whereas the *xz*-plane was the *H*-plane. The radiation pattern in the *E*-plane is observed to be bi-directional whereas that in the *H*-plane is observed to be omni-directional at 1.7375 GHz, as depicted in Fig. 14.

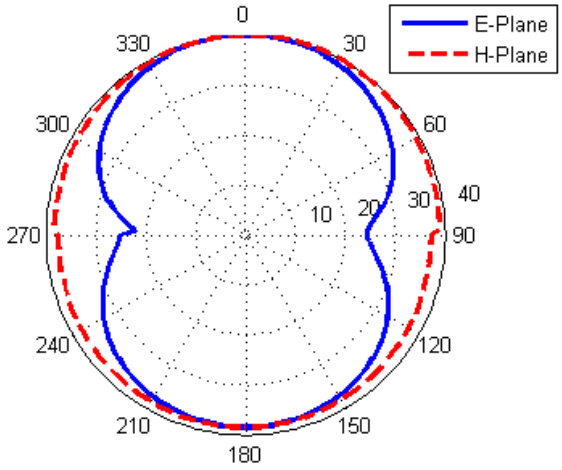


Fig. 14 Radiation pattern of the 1st iteration FACRDGPS at 1.7375 GHz

Similarly, from the current distribution plot (in vector form) of the 1st iteration FACRDGPS at 2.0375 GHz shown in Fig. 13(b), it is observed that current flows in the *y*-direction. Thus, the *yz*-plane is the *E*-plane whereas the *xz*-plane is the *H*-plane. The radiation pattern in the *E*-plane is observed to be bi-directional whereas that in the *H*-plane is observed to be omni-directional at 2.0375 GHz, as depicted in Fig. 15.

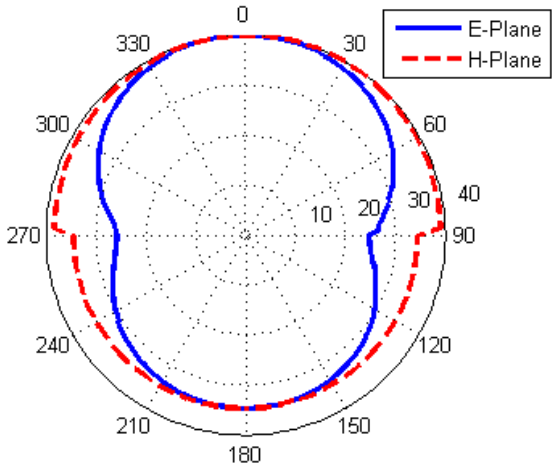


Fig. 15 Radiation pattern of the 1st iteration FACRDGPS at 2,0375 GHz

C. 2nd iteration FACRDGPS

The return loss plots of the 2nd iteration Fractal antenna with concentric ring-shaped DGPS (2nd iteration FACRDGPS) are shown in Fig. 16. The return loss plots of the Concentric Ring-shaped DGPSMA and fractal iteration 2 antenna are included for comparison.

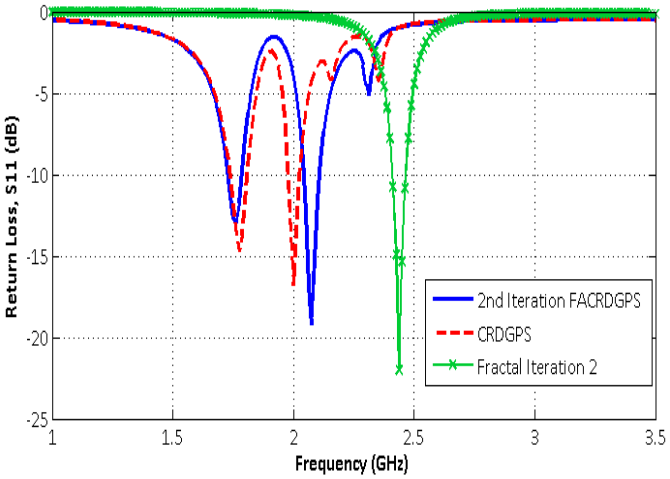


Fig. 16 Combined return loss plot of the 2nd iteration FACRDGPS, Concentric Ring-shaped DGPSMA and the 2nd iteration Fractal antenna

The 2nd iteration FACRDGPS exhibits dual-band resonance, with resonances at 1.7625 GHz and 2.0750 GHz, with –10 dB impedance bandwidths of 58.8 MHz and 66.8 MHz, respectively. The return losses (S_{11}) were -13.0195 dB and -19.2408 dB

The 2nd iteration FACRDGPS current distribution plots are shown in Fig. 17 and Fig. 18 at resonances of 1.7625 GHz and 2.075 GHz, respectively.

From Fig. 17(b) and Fig. 18(b) for the 2nd iteration FACRDGPS, compared to Fig. 9(b) for the conventional patch antenna, it can be observed that the disturbance of the current distribution caused by the excoriation of the fractals in the patch and the concentric ring-shaped defects in the ground plane causes the current to flow in two different paths. This leads to the dual-band resonance observed at 1.7625 GHz and 2.075 GHz.

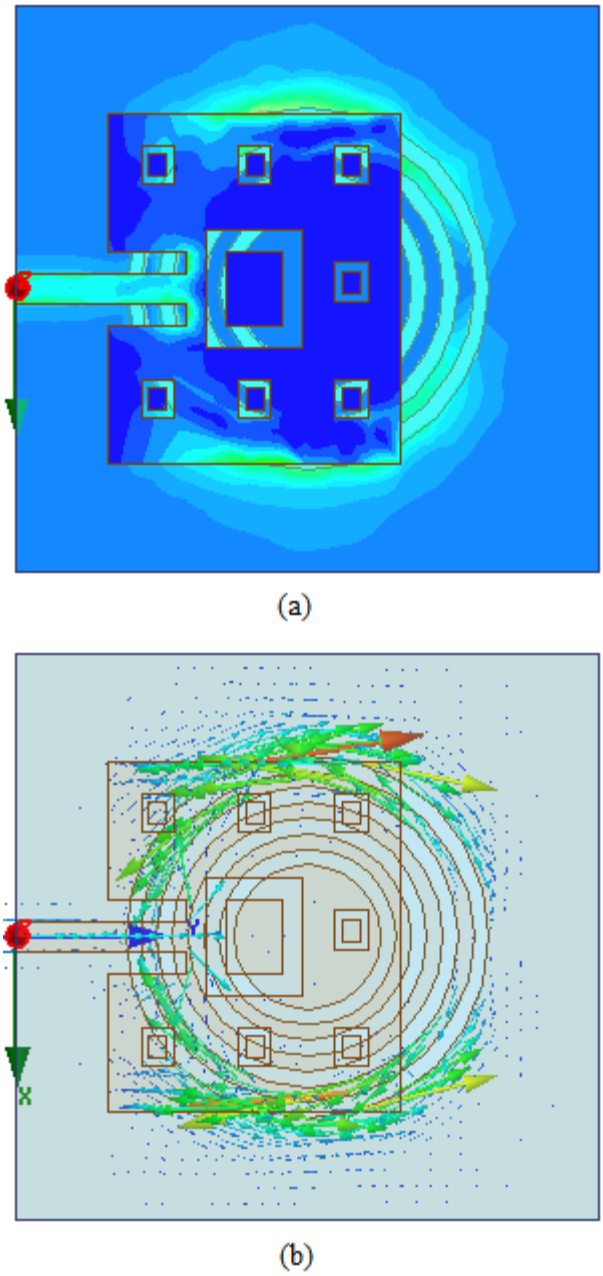
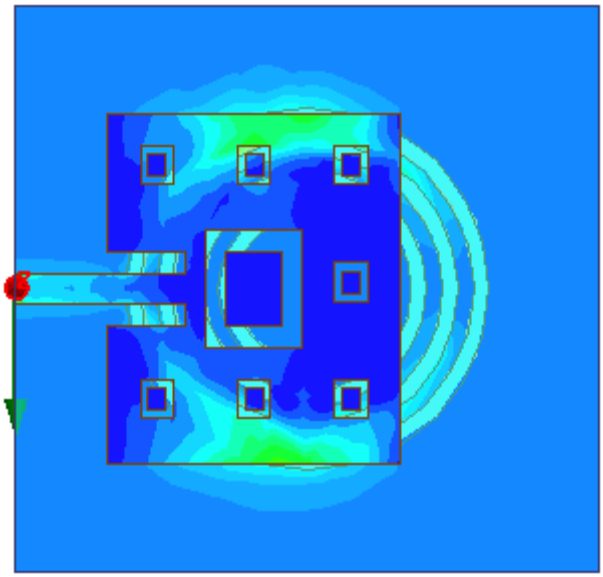
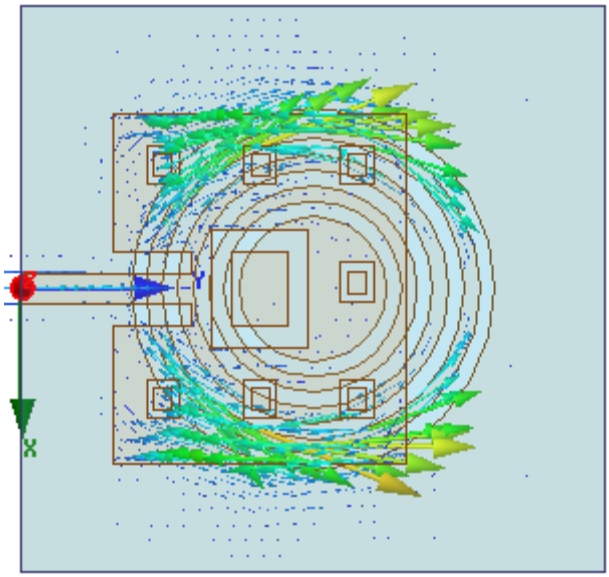


Fig. 17 Current distribution of the 2nd iteration fractal antenna with concentric ring-shaped DGPS (2nd iteration FACRDGPS) at 1.7625 GHz; (a) Magnitude (b) Vector



(a)



(b)

Fig. 18 Current distribution of the 2nd iteration fractal antenna with concentric ring-shaped DGPS (2nd iteration FACRDGPS) at 2.075 GHz; (a) Magnitude (b) Vector

The radiation patterns of the 2nd iteration FACRDGPS are plotted in Fig. 19 and Fig. 20, at 1.7625 GHz and 2.075 GHz, respectively. From the current distribution plot (in vector form) of the 2nd iteration FACRDGPS at 1.7625 GHz as shown in Fig. 17(b), it is observed that current flows in the *y*-direction. Thus, the *yz*-plane is the *E*-plane whereas the *xz*-plane was the *H*-plane. The radiation pattern in the *E*-plane is observed to be bi-directional whereas that in the *H*-plane is observed to be omni-directional at 1.7625 GHz, as depicted in Fig. 19.

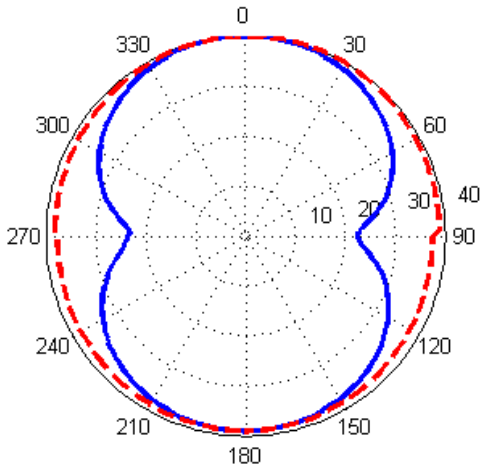


Fig. 19 Radiation pattern of the 2nd iteration FACRDGPS at 1.7625 GHz

Similarly, from the current distribution plot (in vector form) of the 2nd iteration FACRDGPS at 2.075 GHz shown in Fig. 18(b), it is observed that current flows in the *y*-direction. Thus, the *yz*-plane is the *E*-plane whereas the *xz*-plane was the *H*-plane. The radiation pattern in the *E*-plane was observed to be bi-directional whereas that in the *H*-plane was observed to be omni-directional at 2.075 GHz, as depicted in Fig. 20.

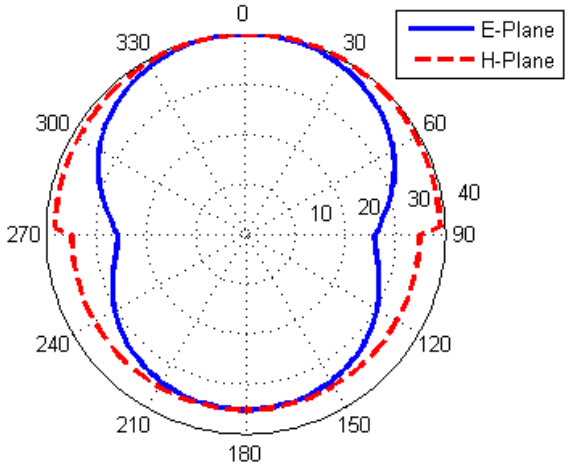


Fig. 20 Radiation pattern of the 2nd iteration FACRDGPS at 2.075 GHz

Fig. 21 shows a combined VSWR plot of the 1st and 2nd iterations of the FACRDGPS. For the 1st iteration, the VSWR is < 2 for the frequencies 1.7011-1.7785 GHz for the lower resonance band and 2.0054-2.0733 GHz for the upper resonance band. For the 2nd iteration, the VSWR is < 2 for the frequencies 1.7235-1.7885 GHz for the lower resonance band and 2.0399-2.1097 GHz for the upper resonance band.

The VSWR being < 2 in these resonance bands indicates good impedance matching.

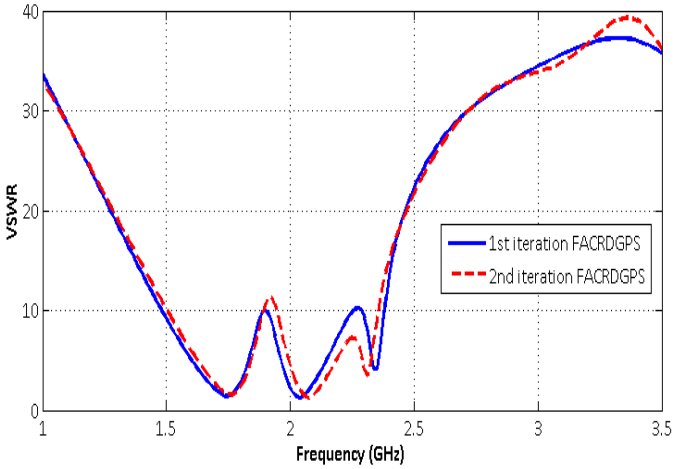


Fig. 21 Combined VSWR plot of the 1st and 2nd iteration FACRDGPS

Thus in summary, from Fig. 11 and Fig. 16, for the 1st and 2nd iteration FACRDGPPS, respectively, it can be observed that incorporating defected ground plane structure in addition to the fractals in the patch in an antenna leads to dual-band performance for both iterations, when compared to the antenna in [20], which only incorporates fractals in the patch and has single band resonance.

D. FABRICATION AND EXPERIMENTAL RESULTS

The 1st iteration Fractal antenna with Concentric Ring-shaped DGPS (FACRDGPS) is fabricated. The 1st iteration FACRDGPS is chosen for fabrication since it exhibits dual-band resonance with larger bandwidth and matching (with regard to lower return loss) when compared to the 2nd iteration FACRDGPS.

Fig. 22 shows images of the fabricated 1st iteration FACRDGPS. The fabrication is undertaken on a pre-sensitized double-sided printed circuit board made of FR4 epoxy substrate with a thickness h=1.6 mm and dielectric constant $\varepsilon_r = 4.7$.

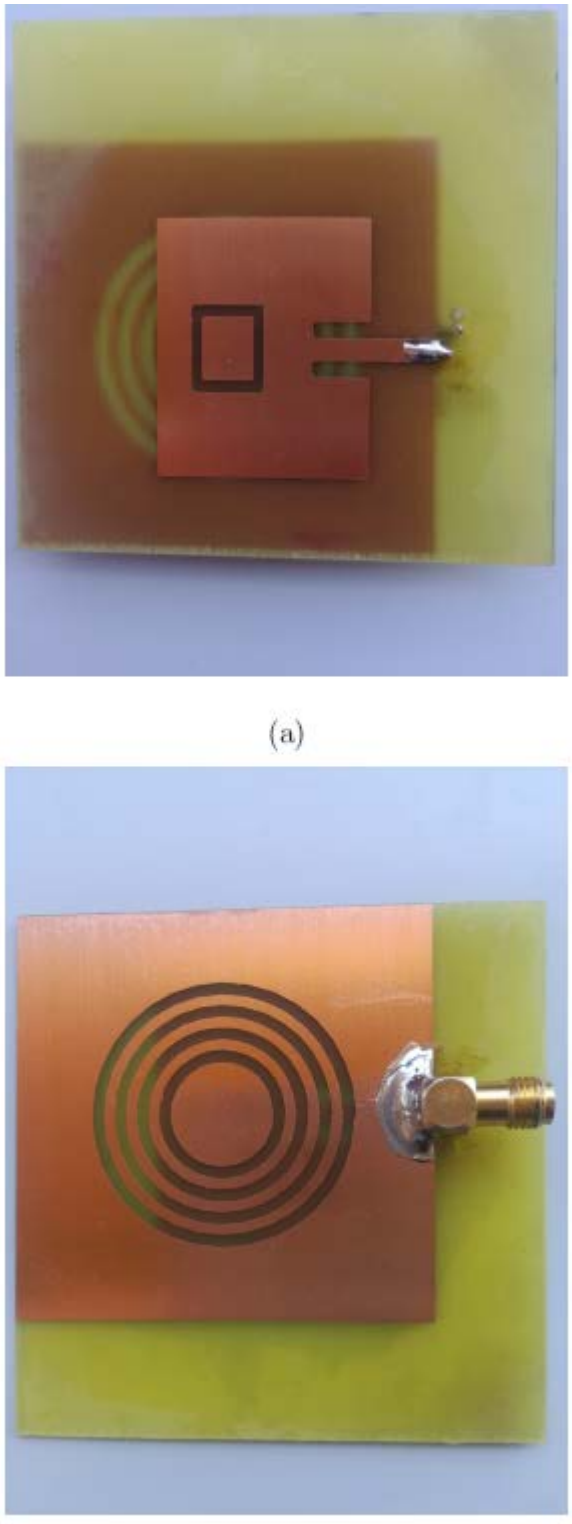


Fig. 22 Images of the fabricated 1st iteration Fractal antenna with Concentric Ring-shaped DGPS (FACRDGPS) (a) Patch (b) Ground plane

The fabricated patch antenna is tested in the laboratory using an MS2027C Anritsu Vector Network Analyzer (VNA) master. Before the return loss measurements are taken, the Anritsu VNA is calibrated for accurate measurement. Since only one

port is used for the return loss (S_{11}) measurement, only Port 1 of the Vector Network Analyzer is calibrated.

The experimental setup is shown in Fig. 23.

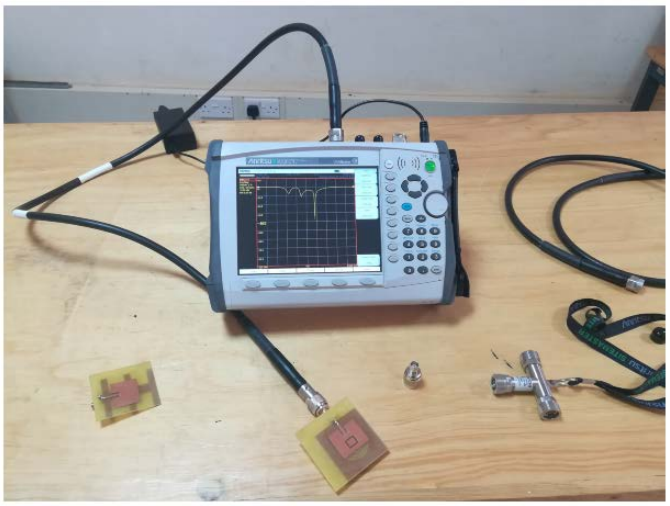
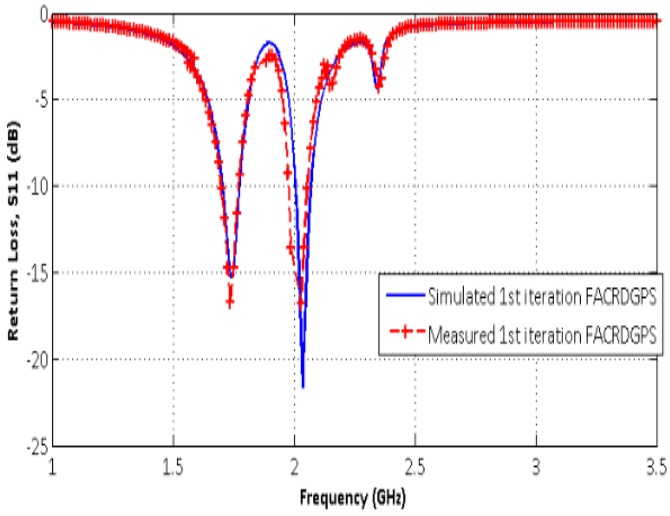


Fig. 23 The experimental set up used for the return loss measurement using the MS2027C Anritsu Vector Network Analyzer

Fig. 24 compares the simulated and the measured return loss plots of the 1st iteration Fractal antenna with Concentric Ring-shaped DGPS (1st iteration FACRDGPS).



From Fig. 24, the fabricated 1st iteration Fractal antenna with Concentric Ring-shaped DGPS (FACRDGPS) exhibits dual-band resonances at 1.74 MHz and 2.03 MHz, with bandwidths of 73.4 MHz and 67.6 MHz, respectively, compared to the simulated bandwidths of 72.5 MHz and 65.5 MHz. Thus there is good agreement between the simulated and the measured return loss. The slight shift in the resonance frequencies can be attributed to tolerance in the fabrication process.

IV. Conclusion

In this paper, dual-band antennas have been designed by combining the fractal design technique on the patch and the defected ground plane structure in the ground plane in an inset fed micro-strip antenna. The fractal design technique only produces a single band of resonance. By incorporating concentric ring-shaped defected ground structures in the ground plane of the same fractal antenna, better antennas with dual-band resonance are designed.

The designed antennas also exhibit low return losses which indicate good impedance matching. For all antennas, the VSWR is also < 2, which also indicates good matching.

The measured return loss of the 1st iteration FACRDGPS also compared well with the simulated return loss.

The designed antennas can be used for mobile communication, and satellite radio applications since the resonance frequencies are located in UHF band.

References

[1] C. A. Balanis, "Microstrip antennas," in *Antenna Theory: Analysis and Design*. New Jersey, United States of America: John Wiley and Sons, 2005, ch. 14, pp. 811-872.

[2] S. Jaiswal, A. Kumar, and N. Kumari, "Development of Wireless Communication Networks: From 1G to 5G," *International Journal of Engineering and Computer Science*, vol. 3, no. 5, pp. 6053-6056, May 2014.

[3] L. Cheema and K. K. Sherdia, "Design of Microstrip Antenna with Defected Ground Structure for UWB Applications," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 2, no. 7, pp. 2525-2528, July 2013.

[4] H. Elftouh et al., "Miniaturized microstrip patch antenna with defected ground structure," *Progress In Electromagnetics Research C*, vol. 55, pp. 25-33, 2014.

[5] T. Wang, Y. -Z. Yin, J. Yang, Y. -L. Zhang, and J. -J. Xie, "Compact Triple-Band Antenna using Defected Ground Structure for WLAN/WIMAX Applications," *Progress in Electromagnetics Research Letters*, vol. 35, pp. 155-164, 2012.

[6] R. Saini and D. Parkash, "Design and Simulation of CPW Fed Slotted Circular Microstrip Antenna with DGS for Wireless Applications," *International Journal of Applied Sciences and Engineering Research*, vol. 3, no. 1, pp. 82-90, 2014.

[7] G. Singh and A. Marwaha, "Design of G-shaped Defected Ground Structure for Bandwidth Enhancement," *International Journal of Computer Applications*, vol. 75, no. 9, pp. 7-11, Aug. 2013.

[8] R. Mudgal and L. Shrivastava, "Microstrip V-slot Patch Antenna using an H-slot Defected Ground Structure," *International Journal of Technology Enhancements and Emerging Engineering*, vol. 2, no. 2, pp. 21-24, 2014.

[9] N. A. Khan and B. A. Singh, "Microstrip Antenna Design with Defected Ground Structure," *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)*, vol. 9, no. 2, pp. 46-50, March 2014.

[10] C. A. Balanis, "Frequency Independent Antennas, Antenna Miniaturization and Fractal Antennas," in *Antenna Theory: Analysis and Design*. New Jersey: John Wiley and Sons, 2005, pp. 611-652.

[11] H. Zang, X. Zhao, X. Xu, G. Zhang, and J. Lu, "A Review of Fractal Antennas," *International Journal of Future Computing and Communication Engineering*, pp. 15-18, 2014.

[12] F. Viani, "Dual-Band Sierpinski Pre-Fractal Antenna for 2.4 GHz WLAN and 800 MHz LTE Wireless Devices," *Progress in Electromagnetics Research C*, vol. 35, pp. 63-71, 2013.

[13] A. Aggarwal and M. V. Kartikeyan, "Pythagoras tree: A fractal patch antenna for multi frequency and ultra-wide bandwidth operations," *Progress In Electromagnetics Research C*, vol. 16, pp. 25-35, 2010.

[14] S. Chauhan, J. K. Deegwal, D. Soni, and P. Singondia, "A Design of Crown-Shape Fractal Patch Antenna," *International Journal of Engineering and Innovative Technology*, vol. 2, no. 3, pp. 177-179, Sep. 2010.

[15] V. Ram, V. Anjaria, P. Boriya, and N. Patel, "Design and Development of Switchable Fractal Patch Antenna for GPS Application," *International Journal of Engineering Science*, vol. 1, no. 7, pp. 46-50, Nov. 2012.

[16] N. Bisht and P. Kumar, "A Dual Band Fractal Circular Microstrip Patch Antenna for C-Band Application," in *Progress in Electromagnetics Research Symposium Proceedings*, Suzhou, China, 2011, pp. 852-855.

[17] M. Susila, T. Rao, and A. Gupta, "A novel smiley fractal antenna (SFA) design and development for ulra-wideband wireless applications," *Progress In Electromagnetics Research C*, vol. 55, pp. 25-33, 2014.

[18] N. Chaudhary, S. Sindhiya, and K. K. TirPhati, "Design and analysis of multiband slotted octagonal fractal antenna," *International Journal of Advanced Research in electronics and Communication Engineering*, vol. 3, no. 1, pp. 41-46, January 2014.

[19] N. Trivedi, S. Gurjar, and A. K. Singh, "Design and Simulation of Novel I Shape Fractal Antenna," *International Journal of Engineering Science and Technology*, vol. 4, no. 11, pp. 4669-4675, Nov. 2012.

[20] S. Shrestha et al., "Design of modified Sierpinski fractal based miniaturized patch antenna," *ICOIN*, pp. 274-279, 2013.

[21] C. A. Balanis, "Microstrip antennas: Feeding methods," in *Antenna Theory: Analysis and Design*. New Jersey, United States of America: John Wiley and Sons, 2005, ch. 14, pp. 813-815.

[22] S. B. Patil, R. D. Kanphade, and V. V. Ratnaparkhi, "Design and Performance Analysis of Inset fed Microstrip Square Patch Antenna for 2.4 GHz Wireless Applications," *IEEE Sponsored 2nd International Conference on Electronics and Communication Systems (ECECS 2015)*, pp. 1194-1200, 2015.

Advances in Friction-stir Processing of Aluminium Alloys

Daniel N. Wang'ombe, Stephen M. Maranga, Bruno R. Mose and Thomas O. Mbuya

***Abstract*—** Friction-stir processing (FSP) is an inspirational approach to the enhancement of metal properties through grain refinement, homogenization of second phase particles, elimination of defects like porosity, and surface modification among others. Although predominantly applied to aluminium alloys, the reach of FSP has now extended to a variety of materials including steels and polymers. This review deals with the fundamental understanding of the process and its metallurgical consequences. The focus is on microstructure, parameters required for microstructure control and mechanical properties of FSP zone; and FSP of selected aluminium alloys.

***Keywords*—** Friction-stir processing, Friction-stir welding, microstructure modification, grain refinement, recrystalization, surface modification.

I. Introduction

METHODS for refining grains in metals include inoculation, control of cooling rate during solidification and severe plastic deformation (SPD). However; inoculation and control of cooling rate produce grain of about 150 μm average size. On the other hand; SPD has been used to develop ultrafine grained (UFG) microstructures ((grain size <1 μm) in aluminium alloys; resulting in improved mechanical properties, such as, strength, ductility, fatigue strength, fracture toughness and superplasticity [1, 2]. Superplasticity is the ability of a material to sustain tensile elongation >200% prior to failure. SPD methods include: equal channel angular pressing (ECAP), high pressure torsion (HPT) and Friction stir process (FSP). UFG material with a submicrocrystalline structure can be produced with only a "single pass" of FSP; As compared with ECAP which generally requires several passes [2].

II. Friction-stir processing

Friction stir processing (FSP) was developed as a generic tool for microstructure modification based on the basic principles of Friction Stir welding (FSW) [3]. Friction stir welding (FSW) was invented in 1991 by Wayne Thomas at The Welding Institute (TWI) in Cambridge, and it has been studied as a solid state welding technique, especially for aluminium alloys [3,4]

Principles of Friction Stir Welding

Friction-stir welding (FSW) is a solid-state, joining process [5, 6, 7] in which a rotating tool with a shoulder and terminating in a pin, moves along the butting or lapping surfaces of two rigidly clamped plates placed on a backing plate as shown in Fig. 1

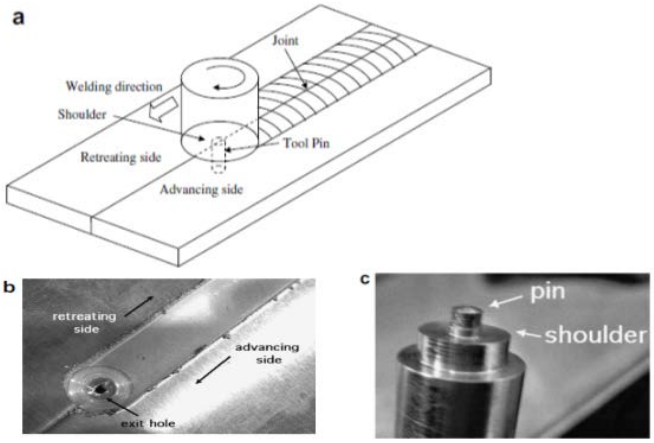


Fig. 1 (a) Schematic illustration of (FSW) process; (b) FSW weld between aluminium sheets; (c) FSW cylindrical tool with a threaded pin [5]

The shoulder makes firm contact with the top surface of the work-piece. Heat generated by friction at the shoulder and to a lesser extent at the pin surface, softens the material being welded. Severe plastic deformation and flow of this plasticised metal occurs as the tool is translated along the welding direction. The material is transported from the front of the tool to the trailing edge where it is forged into a joint.

The half-plate where the direction of rotation is the same as that of welding is called the advancing side, with the other side designated as being the retreating side. This difference can lead to asymmetry in heat transfer, material flow and the properties of the two sides of the weld. For example, the hardness of particular age-hardened aluminium alloys tends to be lower in the heat-affected zone on the retreating side [5].
Once the shoulder makes contact, the adjacent thermally softened region takes up a frustum shape shown schematically in Fig. 2, and corresponds to the overall tool geometry [6]. The region appears much wider at the top surface in contact with the shoulder, tapering down to the pin diameter. A typical cross-section of the FSW joint consists of four zones; namely

F D.N. Wang'ombe, Department of Mechanical Engineering, JKUAT (+2540722283492; e-mail: wangombedanielngera@tum.ac.ke).
S.M. Maranga, Department of Mechanical Engineering, JKUAT (e-mail: smmaranga@yahoo.com).
B.R. Mose, Department of Mechanical Engineering, JKUAT (e-mail: mbruno@eng.jkuat.ac.ke).
T.O. Mbuya, Department of Mechanical and Manufacturing Engineering, UoN (e-mail: tmbuya@uonbi.ac.ke).

base metal (BM), heat-affected zone (HAZ), thermomechanically affected zone (TMAZ), and stirred zone (SZ). The HAZ is similar to that in conventional welds although the maximum peak temperature is significantly less than the solidus temperature, the heat source is rather diffuse, and can lead to somewhat different microstructures when compared with fusion welding processes. The SZ or nugget region containing the ‘‘onion-ring” appearance is the one which experiences the most severe deformation, and is a consequence of the way in which the tool deposits material from the front to the back of the weld. The TMAZ lies between the HAZ and SZ; the grains of the original microstructure are retained in this region, but often in a deformed state. FSW results in microstructural changes which influences the weld mechanical properties [7].

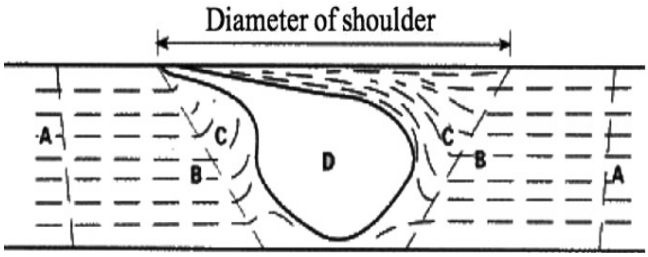


Fig. 2 Schematic cross-section of a typical FSW weld showing four distinct zones: A, B, C and D; namely base metal (BM), heat-affected zone (HAZ), thermomechanically affected zone (TMAZ), and stirred zone (SZ) or nugget or onion zone respectively [5].

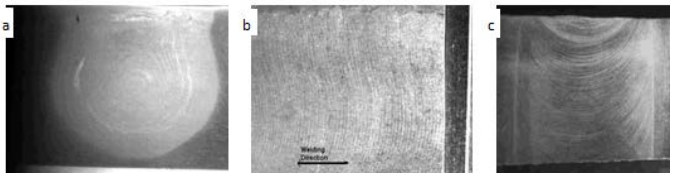


Fig. 3 section views of FSW joint showing the marks at a plane: (a) perpendicular, (b) parallel and (c) traverse to the plane of tool rotation [9].

Various studies have tried to explain the formation of the onion ring in the SZ. The FSW of 2024 to 6061 Al alloys produced intercalated (Swirl and vortex-like) flow patterns in the SZ that were found to be associated with change in tool speed from 400 -1200 rpm [8]. Krishnan [9] found that the 6061 and 7075 FSW joint cross section (a) were characterized by presence of onion rings as shown in Fig. 3. The rings formed swirl patterns in a plane perpendicular to the rotation plane of the tool. Krishnan explained that the formation of onion rings was a geometric effect due to the fact that cylindrical sheets of material were extruded during each rotation of the tool, and cutting through the section of the material produced an apparent ‘Onion Rings’. It was postulated that the tool appeared to wait for a very short time to produce frictional heat and extrude a cylindrical shaped material around to the retreating side of the joint. The spacing of the markings was found to be equal to the forward motion of the tool in one rotation. At a constant feed rate, the spacing of the rings was found to be inversely proportional to the rotation speed.

A recent study [10] on the role of FSW tool on material flow and weld formation in 7020-T6 alloy, found that there were two different modes of material flow regimes involved in the friction stir weld formation; namely “pin-driven flow” and “shoulder-driven flow”. Material transfer in the pin-driven region took place in successive layers that had crescent cross-section with wide top and narrow bottom, and had an upward flow. The shoulder deflected the pin-driven material from the retreating side to advancing side. The layer thickness in the centre of the weld was comparable with distance travelled per rotation of the tool. These material flow regimes merged, and resulted in formation of a defect-free weld. The etching contrast in the regimes gave rise to onion ring pattern in the FSW joint.

FSW joints are known to be free from defects like porosity, slag inclusion, solidification cracks, etc., and these defects deteriorate the weld quality and joint properties. The defects are minimised due to absence of melting during welding, since the metals are joined in the solid state itself due to the heat generated by the friction and flow of metal by the stirring action. However; FSW joints are prone to other defects like pin holes, tunnel defect, piping defect, kissing bond, cracks, among others. The defects result from improper plastic flow and insufficient consolidation of metal in the FSP region [2]. Defects have been observed when the optimal temperature and hydrostatic pressures are not maintained during FSW [9].

Presently, FSW is commercially used in industries, such as ship-building, high-speed train manufacturing, and aviation industry [30].

The principles of FSP

The concept of FSP technique is similar to that of FSW, and is illustrated in Fig. 4. A high strength tool shown in Fig. 4(c), with a shoulder and a pin attached underneath is used for friction stirring. Heat, generated by the friction between the rotating tool and the material surface and the high-strain-rate plastic deformation, softens the material (without melting). The tool traverses through the material, which is then brought from the front to the back of the tool and forged down under the action of the tool shoulder.

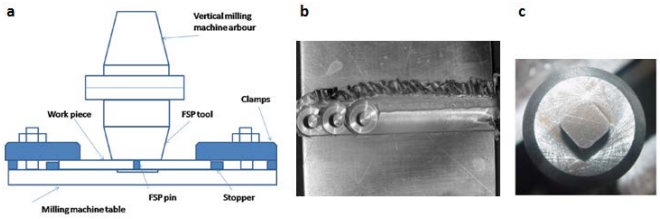


Fig. 4 (a) Experimental set-up of FSP on a milling machine; (b) FSP’ed sample using 3 overlapping passes (c) FSP tool photo showing the tool shoulder and the square pin [11].

Table I Materials and tools used during the FSP/FSW.

Material (mm)				Tool specifications (mm)				Ref
Alloy	L	W	T	Material	Shoulder (D)	Pin (d/square)	length (L)	
7xxx				Cobalt	11.4	Cylinder (4.2)	4	1
6061	300	150	6	High carbon steel	18 (optimum)	4x4		2
1050	100	22.5	5	hardened SKD 61 (JIS)	7	Conical (3)	3	4
7020	300	75	4.4	(HDS) H13 (tool steel)		Frustum(6,4)	4.2	10
6082	120	100	6	MO-W tool steel	15	Square (6X6)	5	11
7xxx		100	7	Ferro-Titanit	18	8	6.8	15
356			6.35					16
6061	60	20	3.1	HSS (SKH9)	12	3	2.8	16

The tool is normally made of high-speed steel, High carbon steel (the tool steel for hot work) [5], polycrystalline cubic boron nitride, and wolfram alloys as shown in Table I. The tool pin shapes include cylindrical (straight, tapered and threaded), triangular and square. There is evidence that FSP tools have been manufactured from end mills that were ground to size on a circular grinder and it was ensured that they conformed to standard sized collets [14]. Based on data shown in Table I the ratio of tool shoulder (D) to pin diameter (d) is 2.5; and the pin diameter is always equal to the plate thickness. However the length of the pin is slightly less than the plate thickness.

Attempts to modify the structure of materials with the FSP method have been carried out on aluminium, copper, magnesium, steels, and also composites with a metallic matrix [17]. There are six areas of research interest using FSP on rolled or milled aluminium plates. Mainly, FSP has been established as a potential grain refinement technique for Al alloys including 6061 [2], 319 [12], 356 [13], and 6082 [11]. Further; FSP has homogenized second phase particles and minimized porosity [13]. It has been suggested that the FSP technique may be applicable to homogenization of MMC materials with inhomogeneous microstructure [4]. Additionally; local modification of material properties, such as: strength, ductility, hardness, fatigue life, fracture toughness and corrosion resistance have been enhanced [16]. Finally; FSP has been used for the creation of composite structures in the surface layer of the material by the introduction of a strange phase, e.g. carbon nanotubes, SiC, Al_2O_3 and SiO_2 particles, and others [17, 18].

III. Microstructure of FSP zone

If Studies on the microstructure of a friction stir processed (FSP) zone, has demonstrated that it consists of fine recrystallized grains resulting from a severe plastic deformation of the magnitude experienced in equal-channel angular pressing (ECAP) and high-pressure torsion (HPT) [4]. Additionally the recrystallized grain sizes in 1050 [2], 1100, 6061 [8] and 7050 [7] aluminium alloys ranged from 2 to 10 μm. These sizes were approximately10 to 100 times smaller than in the original work piece materials [4]. FSP was used to create a microstructure with grain size 0.68 μm in as-cast Al–8.9Zn–2.6Mg–0.09Sc (wt.%) alloy [1]. Microstructure of the FSP zones in 1050 aluminium alloy had recrystallized equiaxed grains of size 0.5- 4 μm [4]. These results suggest that materials with very fine grains, on the scale of microns may be produced through FSP.

The presence of equiaxed and fine grains of 2-10 μm in the SZ of AM60 magnesium alloy was observed. The grain refinement microstructure of the SZ indicated the occurrence of dynamic recrystallization due to concurrent intense plastic deformation and frictional heat. The average grain sizes of the FSP were about 6 μm and 9 μm in the 3,500 rpm and 4,500 rpm friction stir processing trials, respectively.

Grain size measured within the FSP zone ranged between 2.0 and 6.0μm, depending on the processing parameters [1]. Relatively small second-phase particles in 1050 aluminium alloy were homogeneously distributed in the central region of the FSP zone as compared with the unprocessed zone as shown in Fig 5.

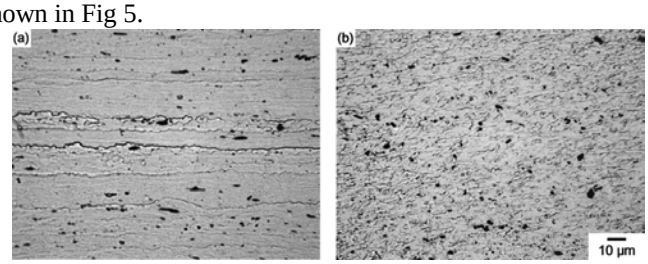


Fig. 5 Comparison of second-phase particles (a) in the unprocessed zone and (b) in the FSP zone of the specimen processed at the tool rotation speed of 560 rpm and the tool traverse speed of 155 mm/min [4].

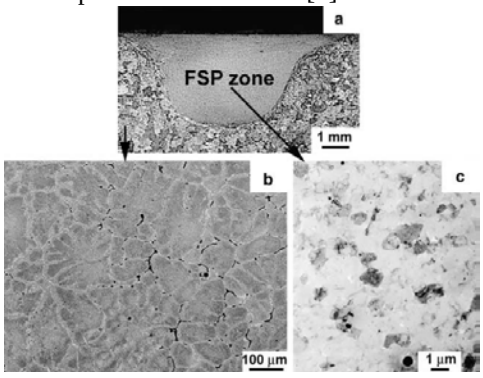


Fig. 6 (a) An optical macrograph of the FSP Al–Zn–Mg–Sc alloy, (b) an optical micrograph of the parent as-cast (unprocessed) material showing typical dendritic microstructure, and (c) a bright field TEM image of the FSP zone, showing ultrafine grains [1].

During the FSP of Al–Zn–Mg–Sc alloy, a combined effect of deformation and temperature lead to very fine grain sizes through the dynamic recrystallization (DRX) process [1,8], as shown in Fig. 6, and break-up of constituent particles [2]. A low peak processing temperature and a fast cooling rate kept the post-DRX growth of grains limited. This, in turn, helped in the generation of the UFG microstructure.

Rolled plates of high strength Al 7050 and Al-Li-Cu alloys have been found to have highly elongated and pancake shaped BM grains [7, 20]. These grains were several hundred micrometers long, approximately 75 μm thick and had average grain diameter of 87 μm. The microstructure of FSW 7050 alloy showed that the BM grains had low angle boundaries, and low dislocation density, and deformation in the TMAZ resulted in severe bending of the grain structure. In contrast, the microstructure within the SZ contained equiaxed grains formed through continuous dynamic recrystallized (CDRX). During CDRX the BM grain low-angle boundaries are replaced by high-angle boundaries in the DXZ by a continuous rotation of the original low-angle boundaries during FSP; under conditions of frictional heating and plastic strain. Consequently; the TMAZ and SZ grains achieves high-misorientation, high dislocation density and high angle grain boundaries. Within the TMAZ, the grain structure of Al 2024 alloy has been reported to be highly elongated and with a considerable grain distortion due to mechanical effect of the tool, and the grain structure of HAZ was observed to be similar that of the BM [21].

Use of a coolant has been found to produce ultrafine grains. Al 6061–T6 plates, with an initial grain size of ∼50μm were subjected to FSP in air and also under water at speed of 1000 rpm and feed of 1.27 mm/s. The final average grain size for air and submerged process were ∼5 μm and ∼200 nm respectively [14]. FSP 7075 Al was quenched by pouring methanol and ice on the joint immediately after the withdrawal of the tool [22]. Microstructural characteristics of five samples from the processed material ∼0, 0.3, 1 and 3mm away from the pin, and the fifth sample was prepared from the joint area outside the tool shoulder were investigated. The average grain sizes starting from specimen 1-5 were 0.1-0.4, 0.8, 1.2, 1.5 and 1-4 μm. Initial sizes of newly recrystallized grains of 7050 alloy were 25–100 nm, and grew to 2–5 μm after 1–4 min due to frictional heating at 350– 450 ºC [23]. Consequently; attempts to minimize the heat generated have prevented the grain growth.

It has been demonstrated that FSP eliminates casting defects like porosity and microsegregation in a single step, and modify the coarser dendritic microstructure into a relatively homogenized one with ultrafine grains (0.68 μm) [1]. The second phase particles have been broken down and distributed homogenously in the SZ [4]. Within the SZ, and due to the intensified stirring action, the second phase particles were homogeneously dispersed and had almost regular size, whereas, on the TMAZ side the particles size is heterogeneous ranging between coarse to fine particles as shown in Fig. 7. In addition, the density of particles within the TMAZ has been found to be less than that in the SZ [11]. Therefore, the microstructural development in each region is usually a strong function of the local thermo-mechanical cycle experienced.

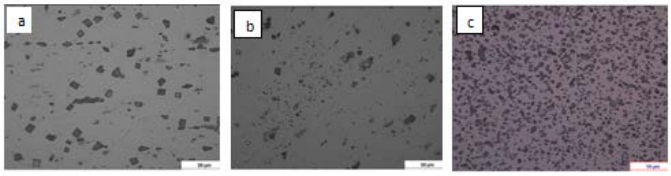


Fig. 7 Second phase particles found in specimen processed at 850 rpm, 90mm/min using one pass taken at different locations: (a) BM, (b) TMAZ and (c) SZ [11].

IV. Parameters Required for Microstructure Control of the FSP Zone

The preparation of UFG Al alloys via FSP has been limited to variation of main parameters including: tool rotational speed, transverse speed (feed rate), the vertical pressure on the tool, the tilt angle of the tool and the tool geometry.

TEM microstructure of the FSP zones in 1050 aluminium alloy at tool rotation speed of 560, 960 and 1840 rpm were composed of recrystallized equiaxed grains 0.5 μm, 1–2 μm and 3–4 μm, respectively [4]. Therefore the recrystallized grain size increased with increase in the maximum temperature from 190 to 310 ºC, which resulted from the increase in the tool rotation speed. The FSW of 2024 to 6061 Al alloys produced a temperature rise that was directly proportional to increase in tool rotation speeds in the range 400–1200 rpm [8]. Consequently; the grains were observed to grow by 10 to 10^2 times the original recrystallized grain. Further peak temperature has been found to increase significantly with increasing rotational speed and vertical pressure on the tool, and to decrease slightly with transverse speed [4]. Therefore the rise in peak temperature has been found to have a marked increase in the grain size. The feed rate increase has been inversely proportional to grain size [19]; however, it has not shown a significant effect on the grain size [1, 3, 4]. Hence; grain growth is observed with an increase in the processing parameters that promote heat generation.

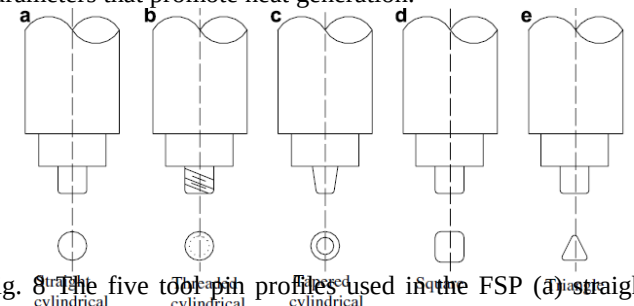


Fig. 8 The five tool pin profiles used in the FSP (a) straight cylindrical, (b) tapered cylindrical (c) threaded cylindrical (d) triangular and (e) square [2].

Tool geometry includes the pin diameter (d), pin length (L), pin profile and shoulder diameter (D). Pin profile has been found to play a crucial role in material flow and in turn regulate the transverse speed of the FSP. During the investigation of the Influence of tool pin profile and tool shoulder diameter on the formation of friction stir processing

zone of 6061 and 2219 alloys [2,24], five different tool pin profiles (straight cylindrical, tapered cylindrical, threaded cylindrical, triangular and square) as shown in Fig. 8 were used, and each profile had three different shoulder diameters.

Eccentricity of the pin allows the hydromechanically incompressible plasticised material to flow more easily from the leading edge to the trailing edge of the rotating tool [6]. Off-centre or non-circular pin allows plasticised material to pass around the pin. The square pin had the highest eccentricity with dynamic volume and static volume in the ratio of 1 to 1.56, and produced the highest stirring pulsating action of 4 pulses per revolution of the tool in the flowing material due to flat surfaces. Consequently, the square profile produced the finest equiaxed grained microstructure, highest tensile strength, greatest microhardness, and least defects. Three shoulder diameters (15, 18 and 21mm) of the square pinned tool were considered. The combined effect of higher eccentricity, higher number of pulsating stirring action during metal flow and an optimum tool shoulder diameter may be the reason for superior tensile properties of the joint fabricated using square pin profiled tool with 18 mm shoulder diameter. Increasing the shoulder diameter has been reported to produce an increase in the heat generated during the FSP, and promoting grain growth [19].

During the investigation of the influence of multi-pass friction stir processing on the properties of alloy 6082, it was found that the accumulated heat accompanying multiple passes resulted in increase in the grain size, dissolution of precipitates and fragmentation of second phase particles [11]. The number of passes had bigger effect than that of the processing parameters.

V. Micro Hardness and Tensile Properties

Vickers microhardness measurements have been conducted across the section of FSW and FSP specimens. Hardness profiles across the SZ, TMAZ, HAZ and BM were found to vary. The microhardness profile of 7005 10% composite shown in Fig. 9 had a minimum hardness in the SZ, accompanied by an increase in the TMAZ to a peak, there after another minima was attained in the interface between TMAZ and HAZ, in the HAZ the hardness increased [15]. However; in another study [11] on FSP of alloy 6082, the lowest microhardness was found in the TMAZ on the advancing side and hardness was found to rise in the BM as shown in Fig. 10 (a). The study failed to distinguish the HAZ from the BM. Additionally from the two studies, it is not clear whether the hardness profile in the TMAZ is increasing or decreasing. Unlike in the hardness of 6082 alloy where the lowest hardness was noted on the advancing side, no significant difference in hardness between the advancing and retreating sides were found in the study of alloy 6061 [16]. The microhardness of the 6061 and 6082 alloys was found to be lowest on the retreating side at the transition between the TMAZ and HAZ, and the point was within the limits of shoulder diameter [25]. In the study of FSW 2024-T351 alloy the SZ was found to be significantly harder than the region of TMAZ immediately outside the SZ boundary [21]. The point of minimal hardness in 5083 alloy has been found [26] to posses high residue stress and tensile fracture occurred at the point in both the 6082 and 6061 alloys [25].

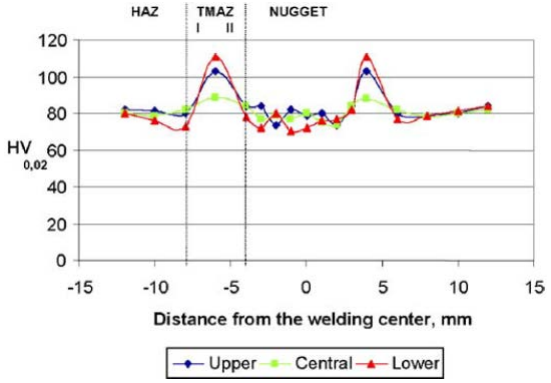


Fig. 9 Hardness profiles on the cross-section of the FSW 7005 10% Al_2O_3 composite [15]

Microhardness measurement results of 6082 and 6061 alloy are as shown in Fig. 10 (a) [11] and Fig 10 (b) [16] respectively. In general, it can be seen that the SZ became much softer than the unaffected BM. The relatively high hardness of the base material was due to heat treatment. Softening of SZ with respect to the BM was due to the decomposition of β" precipitates, over ageing of the precipitates and fragmentation of second phase particles. In addition, the TMAZ experienced more softening than the SZ due to the dissolution and growth of precipitates.

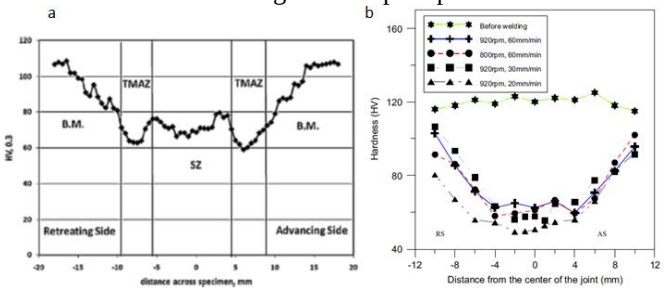


Fig. 10 Hardness distribution across (a) FSP-ed 6082 Al alloy specimen at 850rpm and 140mm/min using two passes [11]; (b) FSW specimen [16].

The microhardness measurements across friction stir welded 2024-T351 joint approximated the lengths of the four distinct zones [21]. The centre of the SZ was significantly harder than the TMAZ immediately outside the SZ, and hardness was lowest in the middle of the TMAZ. There after hardness increased steadily towards the boundary with HAZ, and into the HAZ until the BM hardness was realised. The lengths of the three zones were: the SZ from plate joint line (PJL) extending 5–6 mm out, the TMAZ from 5–6 mm to 10–12 mm, and the HAZ from 10–12 mm to 30–45 mm. The distance from the PJL to the end of the HAZ was approximately 48 mm. The residual hardness of FSW 6061 -T_6 alloy varied from 55 VHN near the top of the weld to 65 VHN near the bottom; in contrast to a workpiece hardness which varied between about 85 and 100 VHN [26]. Microhardness variation across FSW plates of two dissimilar metals shown in Fig. 11b showed a clear distinction between the three zones and the BM [27].

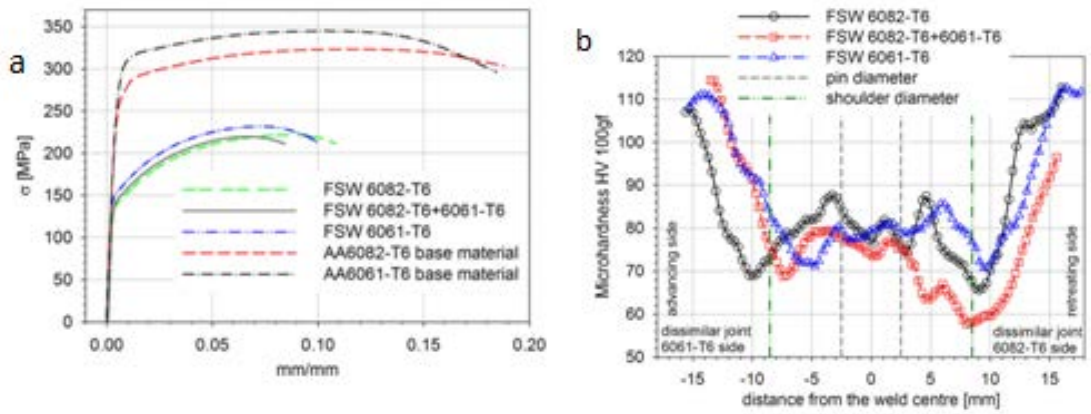


Fig. 11 (a) Tensile tests for FSP specimens and (b) transverse microhardness the specimens [25]

A zone lying in the transition between TMAZ and HAZ on the retreating side had the lowest hardness value [8], and in the case of the mixed joint it was located in the AA6082-T_6 alloy plate side [25].

VI. FSP of Cast Aluminium Silicon Alloys

FSP has been applied to aluminium silicon alloys. Friction stir processing (FSP) was applied to cast 356 Al alloy to modify the as-cast microstructure [13]. FSP eliminated porosity, homogenized and refined the cast microstructure. The resultant microstructure contained fine Si particles (0.25 – 0.42 μm) that were distributed in a fine equiaxed grain aluminium matrix of size 3 – 4 μm. A post-FSP T6 heat treatment did not alter the Si particle distribution, but did significantly coarsen the Si particles. In another study [12] A319 and A356 castings were treated by friction stir processing to reduce porosity and to create more uniform distributions of second-phase particles. The dendritic microstructures were eliminated in stirred zones. The ultimate tensile strengths, ductility, and fatigue lives of both alloys were increased by the FSP.

VII. FSP of 6xxx Series Alloys

The SZ in 6061- T6 and 6082- T6 alloy was characterized by a dynamic continuous recrystallization microstructure, and second-phase particles in the workpiece were essentially "stirred" into the weld zone; resulting in a homogeneous structure containing equiaxed grains and a reduced dislocation density [11, 27]. The SZ grain size averaged 10 μm in contrast to 100 μm for the workpiece. Although TMAZ underwent plastic deformation, recrystallization did not fully occur in this zone due to insufficient deformation strain and lower pick temperatures compared to the SZ. As a result it was characterized by a less deformed structure, in which the parent metal-elongated grains were markedly bent due to plastic deformation into the direction inclined to the TMAZ/SZ boundary. The heat affected zone (HAZ) experienced a thermal cycle, without plastic deformation and thus retaining a grain structure similar to the parent material. The residual hardness of this zone varied from 55 VHN near the top of the weld to 65 VHN near the bottom; in contrast to a workpiece hardness which varied between about 85 and 100 VHN. For the 6061 –T6 and 6082 –T6 alloys it was observed that the minimum microhardness occurred in the transition between TMAZ and HAZ, and on the retreating side [27]. Further tensile failure occurred at the point. An increase of tool rotational speed from 400 to 1200 rpm during the FSW of 2024 to 6061 resulted to rise in temperature to 0.6- 0.8 T_m within the processed zone [8]. Introduction of multi FSP passes on 6082- T6 alloy was found to raise the temperature too [11]. The heat generated dissolved the β'' precipitates, and upon cooling a softer β' precipitate was formed; hence decrease in hardness. Consequently the microhardness was reduced by about 45%, and the grains were observed to grow by 10 to 10^2 times the original recrystallized grain size [11, 16].

FSW of 6061 aluminum alloy to copper was difficult due to the brittle nature of the intermetallic compounds formed in the SZ [25]. The joining of the dissimilar metals 6061 aluminum alloy and copper weld consisted mainly of several intermetallic compounds such as $CuAl_2$, CuAl, and Cu_9Al_4 together with small amounts of α-Al and the saturated solid solution of Al in Cu. The measured peak temperature in the weld zone of the 6061 aluminum side reaches 580 ◦C, which was distinctly higher than the melting points of the Al–Cu eutectic.

VIII. Conclusion

As the concept of FSP is relatively new, some areas need comprehensive investigation to optimize and make it feasible. On the basis of the literature reviewed it has been observed that there is limited research studies on optimizing the process parameters of friction stir processing of 6000 series alloys. This can reduce heat generation, and hence yield the benefits of improved mechanical properties and arrest grain growth after refinement.

Acknowledgement

The authors wish to thank the National Research Fund (NRF) and JKUAT Research, Production & Extension (RPE)

Division, for the financial assistance to the ongoing research on the "*Development of Recycle-friendly Aluminum Alloys for Automotive and Structural Applications''* under grant No JKU/2/4/RP/196.

REFERENCES

[1] Charit I. and MishraLow R.S.; (2005); Temperature superplasticity in a friction-stir-processed; Acta Materialia, Vol. 53; pp. 4211–4223.

[2] Elangovan K. and Balasubramanian V.; (2008a); Influences of tool pin profile and tool shoulder diameter on the formation of friction stir processing zone in AA6061 aluminium alloy; Materials and Design; Vol. 29; pp. 362–373

[3] Sen U. and Sharma K.; (2016); Friction Stir Processing of Aluminum Alloys: A Literature Survey; IJSRSET; Vol. 2; pp. 771-774

[4] Kwon Y. J., Saito N. and Shigema I.; (2002); Friction stir process as a new manufacturing technique of ultrafine grained aluminum alloy; Journal of materials science letters, Vol. 21; pp. 1473 – 1476

[5] Nandan R., DebRoy T., and Bhadeshia H.K.D.H.; (2008); Recent advances in friction-stir welding – Process, weldment structure and properties; Progress in Materials Science, Vol. 53; pp. 980–1023

[6] Thomas W. M. and Nicholas E. D.; (1997); Friction stir welding for the transportation industries; Materials & Design; Vol. 18; Nos. 4 and 6; pp. 269- 273

[7] Su J.-Q., Nelson T.W., Mishra R. and Mahoney M.; (2003); Microstructural investigation of friction stir welded 7050- T651 aluminium; Acta Materialia; Vol. 51; pp. 713–729

[8] Li Y., Murr L.E. and McClure J.C.; (1999) ; Flow visualization and residual microstructures associated with the friction-stir welding of 2024 aluminum to 6061 aluminum; Materials Science and Engineering; Vol.; A271; pp.213–223

[9] Krishnan K.N.; (2002); On the formation of onion rings in friction stir welds; Materials Science and Engineering, Vol. A327; pp. 246–251

[10] Kumar K. and Kailas S. V.; (2008); The role of friction stir welding tool on material flow and weld formation; Materials Science and Engineering, Vol. A 485; pp. 367–374

[11] El-Rayesa M. M. and El-Danafa E. A.; (2012); The influence of multi-pass friction stir processing on the microstructural and mechanical properties of Aluminum Alloy 6082; Journal of Materials Processing Technology; Vol. 212; pp. 1157–1168

[12] Santella M.L., Engstrom T., Storjohann D. and Pan T.Y.; (2005); ''Effects of friction stir processing on mechanical properties of the cast aluminum alloys A319 and A356''; Scripta Materialia, Vol 53, pp 201–206.

[13] Ma Z.Y., Sharma S.R. and Mishra R.S.; (2006); Effect of friction stir processing on the microstructure of cast A356 aluminum; Materials Science and Engineering; Vol. A 433; pp. 269–278

[14] Hofmann D. C. and Vecchio K. S.; (2005); submerged friction stir processing (SFSP): An improved method for creating ultra-fine-grained bulk materials; Materials Science and Engineering; Vol. A 402; pp. 234–241

[15] Ceschini L., Boromei I., Minak G., Morri A. and Tarterini F.; (2007); '' Effect of friction stir welding on microstructure, tensile and fatigue properties of the AA7005/10 vol.%Al2O3p composite'' Composites Science and Technology; Vol. 67; pp. 605–615

[16] Hwang Y. M., Kang Z. W., Chiou Y. C. and Hsu H. H.; (2008); Experimental study on temperature distributions within the workpiece during friction stir welding of aluminum alloys ;International Journal of Machine Tools & Manufacture; Vol. 48; pp. 778–787

[17] Kemal M., Kulekci, Esme U. and Buldum B.; (2015); Critical analysis of friction stir-based manufacturing processes; Int J Adv Manuf Technol

[18] Bahrami M., Dehghani K., Givi M.K.B.; (2013); ''A novel approach to develop aluminium matrix nano-composite employing friction stir welding technique''; Materials and Design.

[19] Commin L., Dumont M., Masse J.-E., and Barrallier L.; (2009); Friction stir welding of AZ31 magnesium alloy rolled sheets: Influence of processing parameters; Acta Materialia; Vol. 57; pp. 326–334

[20] [20] Jata K.V. and Semiatin S.L.; (2000); continuous dynamic recrystallization during friction stir welding of high strength aluminum alloys; Scripta mater.; Vol. 43; pp743–749

[21] Bussu G. And Irving P.E.; (2003); the role of residual stress and heat affected zone properties on fatigue crack propagation in friction stir welded 2024-T351 aluminium joints, International Journal of Fatigue Vol. 25 ; pp. 77–88

[22] Ouyang J., Yarrapareddy E. and Kovacevic R.; (2006); Microstructural evolution in the friction stir welded 6061 aluminum alloy (T6-temper condition) to copper; Journal of Materials Processing Technology; Vol. 172; pp. 110–122

[23] Rhodes C.G., Mahoney M.W., Bingel W.H. and Calabrese M.; (2003); Fine-grain evolution in friction-stir processed 7050 aluminum; Scripta Materialia; Vol. 48; pp. 1451–1455

[24] Elangovan K. and Balasubramanian V.; (2008b); Influences of tool pin profile and welding speed on the formation of friction stir processing zone in AA2219 aluminium alloy; journal of materials processing technology ; Vol. 2 00 ; pp. 163 – 175.

[25] Moreira P.M.G.P., Santos T., Tavares S.M.O., Richter-Trummer V., Vilaça P., and de Castro P.M.S.T.; (2009); Mechanical and metallurgical characterization of friction stir welding joints of AA6061-T6 with AA6082-T6; Materials and Design; Vol 30; pp 180–187

[26] Peel M., Steuwer A., Preuss M., and Withers P.J.; (2003); Microstructure, mechanical properties and residual stresses as a function of welding speed in aluminium AA5083 friction stir welds; Acta Materialia, Vol. 51; pp. 4791–4801

[27] Liu G., Murr L.E., Niou C-S., McClure J.C., and F.R. Vega; (1997); Microstructural aspects of the friction-stir welding of 6061-T_6 aluminum; Scripta Materialia; Vol. 37, pp. 355-361

[28] Su J. -Q, Nelson T. W., and Sterling C. J.; (2005); Microstructure evolution during FSW/FSP of high strength aluminum alloys; Materials Science and Engineering; Vol. A 405; pp. 277–286

[29] Rhodes C.G., Mahoney M.W., Bingel W.H., Spurling R.A. and Bampton C.C.; (1997); Effects of friction stir welding on microstructure of 7075 aluminum; Scripta Materialia; Vol. 36; pp. 69-15.

[30] Cam G. and Mistikoglu S.; (2014) Recent Developments in Friction Stir Welding of Al-alloys; Journal of Materials Engineering and Performance

Review of the Application of Genetic Algorithm and Precision Points in Optimisation of the Four-bar Mechanism

Joseph K. Mwangi, Onesmus M. Muvengei, and Moses F. Oduori

Abstract—Four-bar linkages are the most popular mechanisms and several studies have been dedicated to improving them. One area that has attracted the attention of many researchers is kinematic synthesis. Kinematics synthesis is the reverse of kinematic analysis in that while the latter involves the determination of the motion in a mechanism, the former is concerned with the development of a mechanism to satisfy certain desired motion specifications, either displacement, velocity or acceleration or a combination of any of them. A specific task of kinematic synthesis is of interest: This is the dimensional synthesis which deals with the synthesis of linkage lengths. Over the years, two main techniques have been used for dimensional synthesis: Precision point synthesis and genetic algorithm optimisation. Precision point synthesis was the initial method in the optimisation of the four-bar mechanism designs. Recently, the use of evolutionary techniques such as genetic algorithm (GA) have made it easier to create efficient mechanisms courtesy of advancements in computer technology. The emergence of computational software such as MATLAB simplifies the use of GA. This paper reviews how precision points and genetic algorithms have been used to optimise the four-bar mechanism.

Keywords—dimensional synthesis,four-bar mechanism, genetic algorithm,optimisation, precision points

I. INTRODUCTION

THERE has been great research interest in the field of optimisation of the four-bar mechanism. Mechanism optimisation mostly involves a section of kinematic synthesis known as dimensional synthesis. Dimensional synthesis is the determination of mechanism elements such as the length of the members, angles and coordinates which are necessary for the creation of a mechanism with a desired motion. Researchers on dimensional synthesis have focused on two main areas, namely, the precision point synthesis and application of mathematical programming techniques to mechanism design [1],[2].

Precision point synthesis has excellent properties in motion geometry design optimisation. However, it has limitations in cases where the desired mechanism ought to satisfy certain constraints. In such cases, trial and error bases are used to repeatedly revise the design, a process which becomes very cumbersome. To solve such challenges comes the mathematical programming techniques. A combination of the two approaches is desirable for the best mechanism optimisation results [1]. One of the most reliable mathematical programming techniques is the genetic algorithm(GA).

Joseph K. Mwangi, Department of Mechanical Engineering, JKUAT (e-mail: joseph.mwangi@jkuat.ac.ke).

Onesmus M. Muvengei, Department of Mechanical Engineering, JKUAT (e-mail: mmuvengei@eng.jkuat.ac.ke).

Moses F. Oduori, Department of Mechanical and Manufacturing Engineering , UoN (e-mail: foduori@uonbi.ac.ke).

GA is part of an applied research process known as Evolutionary Computation used for evaluating optimum solutions [3]. As the name suggests, the process is analogous to Darwin's theory of natural selection where only the fittest survives. GA is a non-deterministic process that uses stochastic information to obtain global optimum values. Holland,(1962), [4] was the first to use the GA technique, though its use as an optimisation tool began in the late 1980s [2].

GA possesses exceptionally excellence in handling ill-behaved, discontinuous and non-differentiable algorithm functions. In its operations, unlike the biological genetic systems, the GA encompasses some randomness which guarantees better solutions rather than unfavourable ones. The most important advantages of the GA are its simplicity in implementation, fast convergence times and limited knowledge of solution space required to operate it [5].

II. PRECISION POINTS

Dimensional synthesis has two different but related segments called approximate and exact syntheses. Exact synthesis is only possible for some limited cases of "nice" functions [6]. In most synthesis procedures, the aim is to only approximate the desired solutions. The approximate solutions only match the actual solutions at selected points. These selected points are known as the precision or accuracy points [7].

The target link for the points is mainly the output link and in most cases is the coupler. Precision points are used with the assumption that the design will deviate slightly from the desired function in between the precision points and the deviation is within the acceptable region. The deviation is known as the structural error and is usually between 3-4% [8]. The aim of every mechanism design is to minimise the structural error. However, the error is dependent on the number and spacing of the precision points. The number of points chosen should be such that the structural error generated between the points is minimal. The synthesis of these points is controlled by the number of independent design variables that describe the mechanism [2].

The maximum number of accuracy points required is dependent on the type of dimensional synthesis. There are three forms of dimensional synthesis, namely function, path, motion generation[8],[9]. Fig.1 aids in understanding the equations and maximum precision points required in each of these synthesis processes.

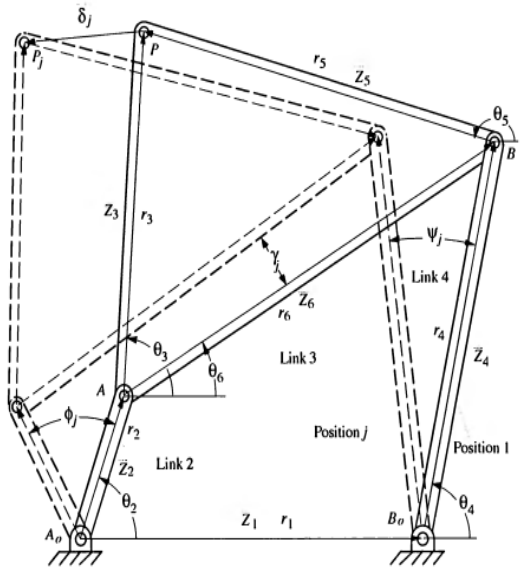


Fig. 1. Vector loop for a four-bar mechanism with a coupler point for two different positions, 1 and j [9]

The vector loop closure equation for position 1 is as shown in (1).

$$\vec{Z_2} + \vec{Z_3} - \vec{Z_5} - \vec{Z_4} - \vec{Z_1} = 0 \tag{1}$$

In position j, the loop is represented by (2).

$$\vec{Z_2}e^{i\phi_j} + \vec{Z_3}e^{i\gamma_j} - \vec{Z_5}e^{i\gamma_j} - \vec{Z_4}e^{i\psi_j} - \vec{Z_1} = 0 \tag{2}$$

where

ϕ_j, γ_j, ψ_j are the angular changes associated with links 2, 3, and 4 respectively.

First, function generation is the correlation of the sliding or rotary motion of the input and output links. For instance, to generate a function $y = f(x)$, the value of x could be related to the crank angle while the value of y could be the output angle at the precision points. From Fig. 1, only one vector is required to represent link 3 during function generation. Therefore, the vector equation for position 1 and j are as shown in (3) and (4) respectively.

$$\vec{Z_2} + \vec{Z_6} - \vec{Z_4} - \vec{Z_1} = 0 \tag{3}$$

$$\vec{Z_2}e^{i\phi_j} + \vec{Z_6}e^{i\gamma_j} - \vec{Z_4}e^{i\psi_j} - \vec{Z_1} = 0 \tag{4}$$

The overall equation for function generation is obtained by subtracting (3) from (4) as shown in (5).

$$\vec{Z_2}e^{i\phi_j - 1} + \vec{Z_6}e^{i\gamma_j - 1} - \vec{Z_4}e^{i\psi_j - 1} = 0 \tag{5}$$

By virtue of being a function generation process, the variables ϕ_j and ψ_j are prescribed. Therefore, the remaining unknowns are the direction and magnitude of vectors $\vec{Z_2}, \vec{Z_6}$, and $\vec{Z_4}$ as well as the angle γ_j. The number of unknowns for n number of precision points can, therefore, be written as $6 + (n - 1)$. If a vector loop equation was to be written for every precision point from point $j = 2$ to $j = n$, the number of scalar equations would be $2(n-1)$ since each vector equation gives rise to two scalar equations. If it happens that the number of unknown variables exceeds the number of scalar equations, there is need to assign values to some variables for deterministic evaluation of the function generation process. Such variables are called free variables or free choices. The number of free choices for function generation is determined from (6).

$$6 + (n - 1) - 2(n - 1) = 7 - n \tag{6}$$

Consequently, the maximum value of n or the maximum number of precision points for function generation is **seven** [9].

Second, path generation involves specifying the path traversed by a point on the output link. The equation for path generation can be obtained by subtracting (1) from (4) while taking note of the definition of the vector for link 3 which is given by (7) and thus obtaining (8).

$$\vec{Z_6} = \vec{Z_3} - \vec{Z_5} \tag{7}$$

$$\vec{Z_2}e^{i\phi_j - 1} + \vec{Z_3}e^{i\gamma_j - 1} - \vec{Z_5}e^{i\gamma_j - 1} - \vec{Z_4}e^{i\Psi_j - 1} = 0 \tag{8}$$

Equation (8) can be rearranged into vector pairs called dyads, as shown in (9) and (10).

$$\vec{Z_2}e^{i\phi_j - 1} + \vec{Z_3}e^{i\gamma_j - 1} = \vec{\delta_j} \tag{9}$$

$$\vec{Z_5}e^{i\gamma_j - 1} + \vec{Z_4}e^{i\Psi_j - 1} = \vec{\delta_j} \tag{10}$$

By virtue of being a path generation process, the displacement vector for point P, $\vec{\delta_j}$, is prescribed. The remaining unknowns are, $\vec{Z_2}, \vec{Z_3}, \vec{Z_4}, \vec{Z_5}, \phi_j, \gamma_j$, and ψ_j. Therefore, the total number of unknown variables for n precision points is $8 + 3(n - 1)$. Meanwhile, there are two vector equations each yielding two scalar equations for every value of n, thus the total number of scalar equations becomes $4(n - 1)$. Consequently, the number of free variables for path generation without prescribed timing is determined as in (11).

$$8 + 3(n - 1) - 4(n - 1) = 9 - n \tag{11}$$

Therefore, the maximum number of precision points required for path generation without prescribed timing is **nine** [9].

However, if there is a correlation between the position of point P and the input link position, the process is called path generation with prescribed timing. In this case, in addition to $\vec{\delta_j}$, either ϕ_j, or Ψ_j are prescribed. Therefore, there is one less unknown and the total number of unknown variables reduces to $8 + 2(n - 1)$. Since the number of scalar equations remain the same, the number of free choices available is given by (12).

$$8 + 2(n - 1) - 4(n - 1) = 10 - 2n \tag{12}$$

Consequently, the maximum number of precision points required for this case is **five**. Since the number of independent

parameters has been reduced by prescribed timing, the synthesis problem requires less precision points and a maximum of five points is enough [9].

Third, motion generation involves guiding a rigid body to move in a specific motion. Usually, the body is attached or lies on the coupler of the mechanism. In this case, the free variables from (4) are: $\vec{Z}_2, \vec{Z}_3, \vec{Z}_4, \vec{Z}_5, \phi_j$, and ψ_j. Both γ_j and δ_j are prescribed. This is, therefore, similar to the path generation with prescribed timing where the total number of unknown variables for n precision points is $8 + 2(n-1)$ whereas the total number of scalar equations is $4(n-1)$. Similarly, (12) applies for the number of free variables and the maximum number of precision points obtained thereafter applies too, i.e. **five**. Despite their similarities in solution, motion generation differs from path generation in that while the former can be described as the motion of a line in a plane, the latter is the motion of a point in a plane [9].

The range of selection for the number of precision points is between two and the maximum number as defined by different forms of dimensional synthesis. If the maximum number of precision points is not used, free variables must be chosen to ensure deterministic solution to the problem. On the other hand, using more precision points than the maximum required leads to non-linearity, difficulties in finding solutions as well as non-feasible solutions, which for instance, contains complex numbers [2].

Typically, precision points at the extreme range of selection are not desirable as they result in very large errors at the middle of the range while having no errors at the extremes. The essence of precision point synthesis is not to match the desired output but to minimize the overall structural error. Nevertheless, some mechanisms demand for the precision of matching the desired output given by the extreme range of precision points. This is the case where a mechanism is to be used to open and close values as controlled by input link parameters. However, for other mechanisms, the best number of precision points is that which gives equalised structural error in the spaces between the precision points [10].

Apart from the number, spacing is also an important part in precision points synthesis. Several techniques have been proposed for determining the spacing of accuracy points. However, the most prominent and efficient technique is the Chebyshev's Spacing method [10]. According to the method, for n number of accuracy points in the range of $x_0 < x < x_f$, the most appropriate spacing for the precision points x_j is given by (13).

$$x_j = \frac{1}{2}(x_0 + x_f) - \frac{1}{2}(x_0 + x_f)cos\left(\frac{\pi(2j-1)}{2n}\right) \qquad (13)$$

The Chebyshev's spacing method, though efficient and convenient, does not result in optimum point spacing. Due to the assumptions made in deriving (13), it is not necessarily true that the method would result in equal minima and maxima of the structural error. The way to achieve optimum design is through trial and error, starting with the Chebyshev's spacing method and plot the resulting errors against the design variable. At points where maximum errors occur, the spacing between points is reduced and the process repeated until the maxima and minimal of structural error are equalized in the entire range of accuracy points [6].

A. Application of Precision point to Four-bar Mechanism Optimisation

Several authors have done research on precision points synthesis. Freudenstein [11], acknowledged as the father of modern kinematics, presented the synthesis of a four-bar linkage using two types of approximation methods. The first method was based on coinciding the ideal and actual function at several precision points. Approximation equations were developed for three different cases namely, three point, four point, and five point case. The author considered the five point case as the limiting order of approximate synthesis for complex functions using the algebraic method. However, Freudenstein insinuated on the possibility of higher order approximations for simple functions. In the second method of approximation, the author made the ideal and actual functions to coincide at only one precision point but also found out that several finite number of derivatives could be made to coincide at the same point up to the desired order. The limitation of this study is that the methodology used requires deep knowledge in algebraic mathematics. In addition, the algebraic method used has a limit of the number of precision points that can be used for synthesis.

McLarnan [12] sought to improve on the idea of Freudenstein. To do this, the author would minimise the structural error by using half as many precision points as used by Freudenstein thus saving on computation time and space. For a mechanism which allowed the synthesis of n points, McLarnan used fewer points by extending the idea of derivative approximation synthesis to not only one point but also to all other interior precision points. Despite the obvious improvements in utilising precision points, the methodology involved trial and error to minimise the structural error making the procedure cumbersome and time consuming. In addition, the analytical method used cannot allow for some of the synthesis tasks that demand the inclusion of many precision points.

Maclaine [13] used a numerical method to perform the Chebyshev optimum linkage design. The research involved an initial approximation with five precision points followed by the application of Newton's method to perform iterations that generated a crank-rocker linkage for solar declination. Th resultant structural error was highly minimised. The gap identified in this study is that it demands the use of an extra method to generate the initial approximation (of the structural error and derivatives) required for the Newton's method. This extra method could either be graphical(which is prone to human errors) or algebraic (which requires extensive algebraic knowledge). In addition, if these two methods did not give equal minimum structural error, the process had to be repeated, which made the exercise cumbersome and time-intensive.

Rose and Sandor [14] used a rather unconventional method for acquiring optimal dimensions of a four-bar mechanism via function generation by departing from the usual manual iteration method. Normally, the number of precision points is

chosen by the designer while the spacing is obtained using the Chebyshev's theorem. After analysing the structural error, the points are then respaced through Freudenstein's respacing formula to minimise the errors with the process being repeated until the errors between the accuracy points are equal. However, this analytical study suggested that the synthesis could be done in one step by forcing the errors between the points to be equal and minimal by applying additional constraints. The minimised errors were obtained using the Newton-Raphson method implemented using a computer programming software. After an initial approximation, the method linearized the non-linear equations using Taylor expansion method and solved for the error. The error was then fed to the initial approximation to get the second approximation. The computational process proceeded until the an equalised extrema was achieved. Nevertheless, this method is not foolproof as an equalised extrema does not necessarily mean minima of error.

Kinzel et al. [15] proposed the use of Geometric Constraint Programming (GCP) in kinematic synthesis for finitely separated precision points. The GCP method used the sketching window of commercial Computer Aided Design (CAD) software to parametrically develop kinematic diagrams. Numerical solvers were also integrated into the software to solve the synthesis equations thus combining the graphical and analytical techniques for better synthesis results and experience. To show the robustness and versatility of the GCP method, the authors used it to demonstrate motion generation for five precision points, path generation for nine precision points and function generation for four precision points. However, GCP has the limitation that only one solution is obtained; the solution nearest to the one sketched by the designer. This could be a local solution and thus does little in searching for a global solution [16].

Modern precision point synthesis problems are making use of advanced computational resources available. Martinez-Alfaro [17] studied the problem of four mechanism synthesis using Simulated Annealing method. This method enabled the author to transform the synthesis problem into an optimisation one and in so doing permitting the use of more precision points than the classical methods (analytical and graphical). In addition, the method also allowed the inclusion of more constraints for the synthesis of a more practical mechanism. The results obtained showed a good control of the trajectory of the mechanism. It is therefore a continuous path generation method instead of a point-to-point path generation process.

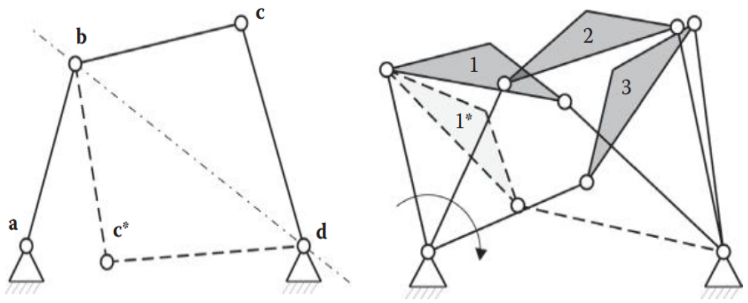
Mehar et al. [18] optimised the dimensions of a four-bar mechanism using five precision points by employing the Freudenstein's equation shown in (14).

$$K_1 cos(\varphi_j) - K_2 cos(\Psi_j) + K_3 = cos(\varphi_j - \Psi_j) \qquad (14)$$

where $K_1 = \frac{a}{d}$; $K_2 = \frac{a}{b}$; $K_3 = \frac{a^2+b^2-c^2+d^2}{2bd}$, φ_j = input angle at point j; Ψ_j = output angle at point j; j = number of precision points; a, b, c, d = frame, crank, coupler, and rocker link lengths respectively. The Least Square Method was used to minimise the structural error with the results being compared to those obtained using four accuracy points. The authors results were more accurate compared to similar results from literature using four precision points showing that the Least Square Method proved is an effective method of minimising structural errors.

B. Limitations of the Application of Precision Points in Four-bar Optimisation

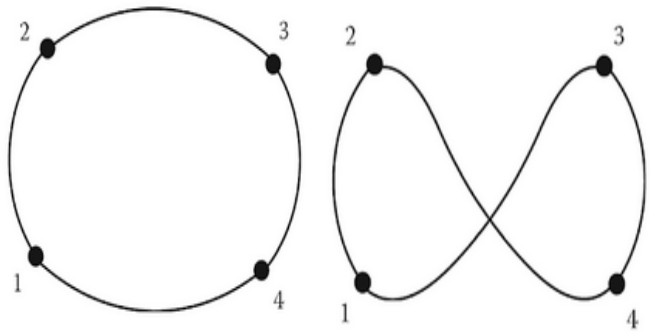
One of the major problems encountered with the application of precision points is branching. Branching occurs when it is physically impossible to have the coupler pass though all the intended precision points without having to be disassemble and reassemble the mechanism. This could happen even though synthesis results show that the coupler is capable of satisfying all the precision points[19],[20]. This occurrence is due to having more than one possible values of the position of a coupler corresponding to a given crank position. The number of mechanism branches is equal to the number of possibilities. In the case of all types of Grashof linkages, there are always two disconnected branches. Fig.2a shows two possible configurations of a four-bar mechanism. From Fig. 2a, a-b-c-d achieves the positions 1-2-3 in Fig. 2b while a-b-c*-d achieves position 1*. Since the mechanism cannot achieve all these positions continuously without having to be disassembled and reassembled, it is said to have a branch defect.



(a) two possible configurations of a four bar mechanism (b) branch defect

Fig. 2. Branching in four-bar mechanism[20]

The second problem in precision points is the order defect. During synthesis, it is normally desirable that the mechanism passes through the precision points in the prescribed order. Although this could happen, the physical outcome of the mechanism might not pass through the points in the specified order [21]. This is illustrated in Fig. 3a and Fig. 3b.



(a) point order 1-2-3-4-1 (b) point order 1-2-4-3-1

Fig. 3. Different path for the same precision points[20]

In graphical synthesis, the branching and order defects can be eliminated using the methods of Filemon's construction [22] and Waldron's construction [23]. However, in analytical synthesis, additional constraint equations as well as prescribed timing help in achieving defect-free mechanisms. Gupta [24] also introduced another method of separating the mechanism branches by partially differentiating the generator function with respect to the output(ϕ).

III. Genetic Algorithms

Genetic Algorithm, or simply GA, as defined by Holland [25], is a search technique that involves proceeding from an initial population of chromosomes (design variables) to a new population via the natural selection and other genetic operators such as crossover, mutation, and inversion. In the recent past, GA has come in handy to solve optimisation problems in the field of engineering giving rise to fresh research ideas as well as applications. This is because most computational problems in this field involve a search through vast amount of possible solutions. In this case, the GA technique offers great versatility by providing a parallel search for solutions, which makes it easier and faster to arrive at the most efficient solution. In addition, GA offers an intelligent approach to choosing the next population to consider for subsequent iterations, which is often desirable [26]. Michalewicz [27] summarised the five fundamental components of GA as follows:

1) A genetic model of the potential solutions to the problem.
2) Generation of an initial population
3) A fitness function to rate the solutions
4) Genetic operators(e.g. mutation operator and crossover operator) to change the offspring composition during reproduction.
5) Values for the GA parameters such as population size, probability of mutation, and probability of crossover.

A. How GA works

It is important to understand how the GA works. The flow chart in Fig. 4 is a summary of the stages of implementation of the GA.

The GA works to optimise a given function $f(x)$ which must be defined beforehand [27]. This is called the objective function. The next stage is the encoding stage where the design variables are represented genetically using an encoding system. For instance, if a binary bit encoding system is used, a single design variable is represented by a gene g_i made up of l_i bits [26]. The number of bits representing the design variable is dependent upon the precision required. A collection of the genes then form a chromosome.

Afterwards, it is necessary to establish the initial population. An initial population is composed of a number of chromosomes that form the first search space. Typically, this is done randomly to avoid biased solutions that tend to fall around particular areas of the search space.

Next stage is the evaluation stage where each chromosome is evaluated according to the fitness function and the current population. The chromosome with the best fitness function value is then stored by the logarithm and becomes a candidate for selecting into the future population.

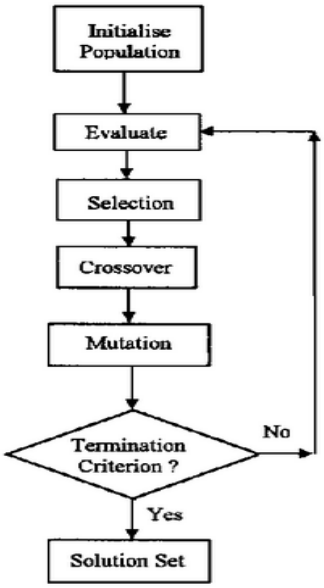


Fig. 4. Stages of the GA process [28]

Taking a step further is a selection process whereby, the current population saved after evaluation becomes the future population. In most cases, the selection process is done through a roulette wheel. The selected chromosomes then become parents for the next generation [28].

Next is the crossover genetic operation in which parts of the parent chromosomes on the different positions of the chromosomes are interchanged to form a different chromosome for the offspring. Crossover can either be single-point, double-point or multi-point. In single-point, there is one crossover point for the parents's genes whereas in two-point crossover, there are two crossover points and so on [29]. It should be noted that crossover occurs for the probability, Crossover Probability, CP $\epsilon[0, 1]$ [5].

There is also the mutation genetic operation which is achieved by random creation of a potential value to replace the old value in the chromosome. It occurs when the operator randomly selects a value in the region of real values $(x_i, x_i \pm range)$ which is then subtracted or added to x_i depending on the direction of mutation. Like crossover, mutation has a probability, mutation probability, (MP) between 0 and 1 but is always less than CP [5]. Both crossover and mutation operations are the two main stages in genetic algorithm.

GA operations can be summed in two phrases: search for the best solution and exploration of the solution space. The genetic operators conduct blind searches and should be directed by the selection operators into feasible areas of the search space [30]. Consequently, all GA elements should be carefully studied and additional heuristics added to improve GA performance.

B. Application of Genetic Algorithm to Four-bar Mechanism Optimisation

Kunjur and Krishnamurty [28] developed a genetic algorithm approach into the synthesis of a crank-rocker mechanism. The synthesis in question was the coupler path generation with prescribed timing. Real number encoding was used to represent the population values. The objective function was an expression of minimization of the structural error which was calculated as a sum of the square of errors at each target point. Constraints used were the Grashof's condition for a crank-rocker and link lengths. Two cases were considered, namely five target point case and an eighteen target point case. From the results, the author's concluded that real number coding achieved better global optima at even a faster rate compared to binary representation. In addition, the eighteen point case yielded more accurate results but required more computational costs in terms of both processor and time. The results were validated against those obtained from two gradient based methods. It emerged that the results were inferior to the exact gradient method results but superior to those of the central difference gradient method. However, the author pointed out that the exact gradient method had its own disadvantages which affect its application in complex mechanism optimisation problems. Nevertheless, this study had some limitations. First, crank angles were considered up to 90° for the five point case and up to 140° for the eighteen point case. Consequently, the entire motion path generation is not complete since one revolution of the crank must be 360°. Second, the target points were created at the user's discretion and were not optimally synthesised as required by the Chebyshev's spacing method or any other optimal space generation technique.

Cabrera et al. [5] also studied genetic algorithm as a solution method for optimal synthesis of planar mechanisms. The method was comparable to that of Kunjur and Krishnamurty [28] albeit with modifications. First, more constraints were included such as the input angle variations and sequence of input angles. Furthermore, the constraints were incorporated into the objective function as penalty functions. Population size and probabilities of the genetic operators were also among the modifications. A disturbing vector was also added to guide the selection operator. Path generation with and without prescribed timing was considered for aligned and misaligned target points. From the study, the authors concluded that GA required more iterations than gradient based methods though there is the benefit of simple function evaluation. The inclusion of additional constraints in this study bore more accurate results in comparison with [28]. However, the methodology in this study could be improved by employing a better technique of creating the precision points instead of just the aligned and misaligned criteria. In addition, some cases considered do not account for the whole range of the input angle.

Dulger et al. [31] presented optimisation of four-bar mechanism using GA in Matlab software(Optimisation Toolbox) and compared the results to those obtained using $fmincon$ in the same software. The type of synthesis was path generation without timing and it utilised six target points arranged in a straight vertical line. After carrying out kinematic analysis of the mechanism, the objective function established was similar to that of Carbrera et al. [5]. According to the results, the use Optimisation Toolbox enhanced faster convergence of the algorithm. For every design variable, the $fmincon$ command yielded only one result while the GA yielded a result for every precision point used. This meant that GA was superior and provided enough data for error analysis. In addition, the Toolbox was easy to use and did not require vast knowledge to perform the optimisation. However, very few mechanisms, if any are known to trace a vertical line of coupler motion and therefore, the results of this study can barely be compared to real-life mechanism application. Furthermore, the study lacked a proper technique to select precision points(number and position) for practical application of the four-bar mechanism synthesis results.

Laribi et al. [32] used a different technique for path generation in a four-bar mechanism. The genetic algorithm approach was augmented with the fuzzy logic method controller whose work was to monitor the variable evolution in the first run and modify the initially set boundaries in readiness for the second run. This adjustment yielded better results in subsequent runs. The flowchart of the Genetic Algorithm-Fuzzy Logic(GA-FL) optimisation was as shown in Fig. 5. The symbols in the flowchart were defined as follows: G = generation; $G_(max)$= maximum number of generations; E_s = structural error; ϵ = tolerance.

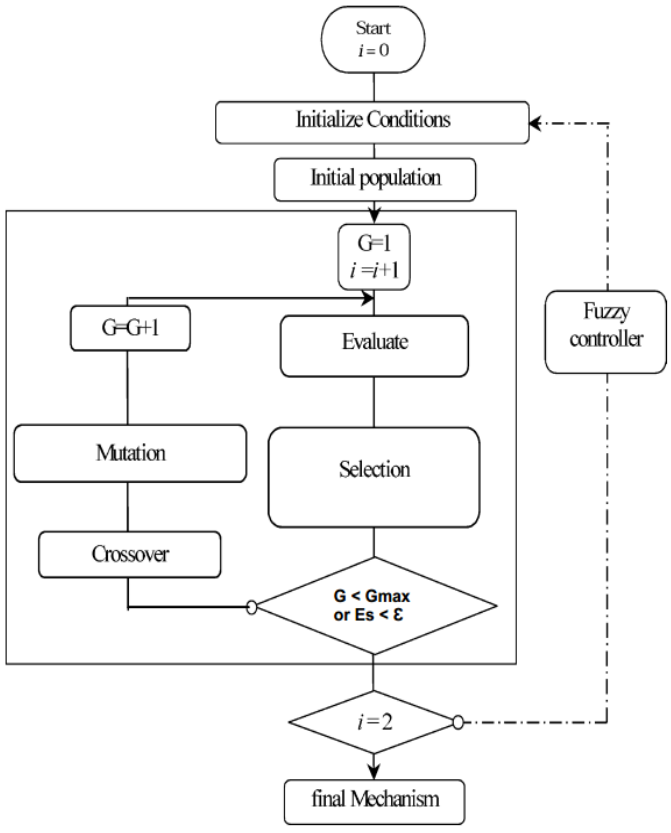


Fig. 5. Optimisation using a GA-FL [32]

Because of the controller, the GA was applied twice in the

optimisation cycle. However, the number of function evaluations were comparable to pure GA methods since convergence was achieved faster. Using this approach, it was possible to reduce the population size required to about half the number in literature without affecting the accuracy of the results. The authors suggested the GA-FL approach was suitable for large domains where GA gave less accuracy compared to deterministic methods. Nevertheless, this method has its shortcomings especially in practical application of the four-bar mechanism whereby manipulating the domain of the design variables could result in a mechanism which does not fit in its assigned location.

Starosta [33] applied GA to synthesis a four-bar linkage as a closed curve generator with the curve being represented by normalised Fourier coefficients (NFCs). By normalising the curve, the number of curve parameters being investigated were minimised resulting in faster convergence of the GA.The objective function was the normal distance between the NFCs and the generated curve. After the synthesis, geometric transformation procedures namely rotation, scaling and reflection were done to determine the configuration of the four-bar linkage. Though the methodology in this paper achieved accurate results, it required the additional procedure of performing transformations after the GA work is complete.

Lin [34] studied path synthesis of the four-bar linkage using a hybrid methodology consisting of GA and Differential Evolution (DE). Though the procedure was majorly similar to that of GA, the aspect of DE was introduced in the crossover stage whereby a differential vector pertubation replaced the normal crossover operation. Consequently, the differential vector became the main perturbation while the mutation operation served as the minor perturbation. As in most other studies, the objective function was the sum of the square of the Euclidean distances between the desired and actual points. This study also established the stopping criteria as the mean error in Euclidean distance. Nevertheless, the author admitted that the proposed hybrid method could not be judged to be better than other evolutionary algorithms and proposed for proper combination amongst evolutionary algorithms for better optimisation potential.

Nariman-Zadeh et al. [35] presented a multi-objective GA application for Pareto optimum four-bar linkage synthesis. The multi-objective aspects came from the attempt to simultaneously minimize two objective functions, namely the tracking error (TE) and the deviation of the transmission angle from 90° (TA). The authors noticed that most synthesis problems were single objective problems based on the tracking errors between actual and desired target points. This was opposed to the real-life optimisation problem which the authors described as multi-objective thus the addition of the TA objective in the study. Pareto fronts showed the existence of two four-bar mechanisms which were a trade-off between the two conflicting objectives under consideration. Nevertheless, even with the benefits of a compromised solution, the problem of definition of transmission angle could render the algorithm inapplicable in some types of four-bar mechanisms whose efficiency is not suitably measured using the transmission angle. This includes such mechanisms as the jaw crusher whose output link is the coupler and not the rocker. Therefore, another objective should replace the TA objective if such mechanisms are to be optimised using the proposed method.

In the work of Acharyya and Mandal [36], the effectiveness of GA in four-bar linkage synthesis was compared to that of two other evolutionary algorithms, namely Particle Swarm Optimisation (PSO) and DE. To do this, the study used the same data and objective as Cabrera et al. [5] but included an additional refinement for selecting the initial population. When the results for the different cases considered were analysed, the DE technique emerged the most accurate followed by the PSO technique. GA was the least accurate with the additional disadvantage of longest convergence time. However, the authors did not clarify why the GA results obtained in the study compared poorly to those obtained by Cabrera et al. [5]. This could be proof that there were some inaccuracies in the preparation of the GA technique in this study.

Chaudhary and Chaudhary [37] utilised GA for two types of optimisation: dynamic balancing and shape optimisation. In dynamic optimisation, the problem involved reducing both shaking forces and shaking moments by applying GA and converting the links to equimomental point-mass systems. Using GA in Matlab, a reduction of 68% and 50% was achieved for shaking moment and shaking force respectively. After dynamic balancing, shape optimisation was done. The link shapes were modelled as cubic B-spline curves. The boundary of the curves were then optimised by considering minimum error in the inertia of the balanced links. Shape optimisation was then done using GA in Matlab and optimum link shapes obtained. However, the results obtained from the study could be erroneous since the problems were not solved to full dimensions as some assumptions were made to truncate the problem size. In addition, the weighted factors assigned to shaking forces and shaking moments were arbitrary hence the probability of error propagation in the optimisation process.

Shinde at al. [38] presented a dimensional synthesis study using GA for a six-bar mechanism which comprised of an input link, fixed link, triangular coupler and output link. The study was an improvement of precision point mechanism synthesis via the incorporation of GA for optimisation. The authors considered two cases of path generation with prescribed timing whereby one case involved an open curve generation with five points while the other involved a closed curve generation with eighteen points. In addition, the study dealt with straight line path generation without prescribed timing with six precision points. The objective function was the structural error calculated as the square of the Euclidean distances between the actual and desired points. After formulation, the objective function was coded into MATLAB and used to solve the three cases. From the resultant plots obtained, the study was successful in attaining paths close enough to those projected. Nevertheless, the task was void of additional constraints that would help to prevent the branching defect. In addition, the errors in the study, though small, could have been further reduced by initially determining the optimum number of precision points required for the different cases.

C. Advantages of the use Genetic Algorithm over Conventional Optimisation techniques in Optimisation Four-bar

Unlike conventional methods which deal with a single sample, the GA searches amongst a population of samples. This reduces the probability of locating false peaks which correspond to false solutions [39]. Therefore, four-bar synthesis results obtained through GA are like to be more accurate.

Moreover, GA uses information about the objective function itself rather than derivative or auxiliary information. Most sequential search methods, for instance, the gradient technique, demands function derivatives in order to carry out search procedures. The fact that GA requires little knowledge of problem structure and parameters is desirable and makes the algorithm more flexible to deal with non-linear, non-differentiable and noisy functions [39]. Some of the constraints equations and variable relationship in four-bar mechanisms are non-linear thus making GA a good candidate technique for synthesis.

Unlike traditional methods which are mostly deterministic, GA possesses stochastic operators that use probabilistic transition rules in the search operation. The organised random search is an important aspect that is aimed at guiding the search into regions of the solution space with better solutions. This makes GA more robust and effective [40].

In addition, while conventional methods act directly on the values of the population, the GA uses a coding system of bit strings. This enables GA to easily handle a variety of design variables [41].

Finally, GA applies multiple operators to manipulate the entire population at each generation thus making the search global and multi-directional. The search space is therefore well exploited and the possibility of being trapped in local optima is low [41]. Moreover, the fact that GA preserves a population of potential solutions reduces the risk of local convergence and promoting the evaluation of global optimum which is good for multi-modal optimisation problems [42].

D. Limitations of GA

Problems associated with GA could be due to the random sampling in population initialisation. When the population size is small, randomisation does not result in satisfactory sampling of the search space as shown in Fig.6a. Preferably, the solution space could be subdivided into several zones and each of the zones thereafter represented in the initial population so as to ensure uniform distribution of the population as shown in Fig 6b. Another approach is to use prior knowledge to bias the solution space towards a region rich in potential optimum solutions as shown in Fig 6c. This may be helpful in getting to the global optima faster as well as getting high quality results without having to increase the population size [27],[41].

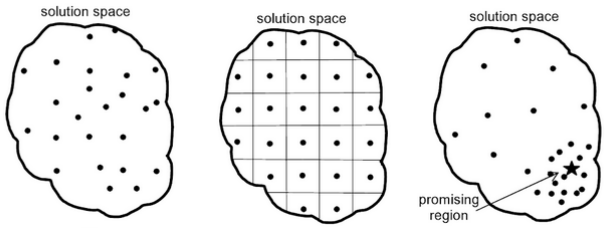


Fig. 6. Population Initialisation approaches [41]

Another limitation of GA is infeasibility and illegality. It should be understood that GA search operations involve alternating between the coding space and the solution space. The coding space is the genotype space while the solution space is the phenotype space. Usually, the GA operators function in the genotype space while the evaluation and selection occur in the phenotype space [30]. The transformation from the genotype to the phenotype is an important feature in GA. Sometimes, the transformation could yield results which are infeasible to the optimisation problem especially in cases of constrained problems and multi-modal optimisation [30]. Infeasible solutions are those that are within the solution space but are outside the expected feasible region. In other cases, illegal solutions could result from the GA. Illegal solutions are those that when decoded do not even fall within the solution space. This includes results consisting of complex numbers. Practically, the lengths of the four-bar links cannot be complex numbers meaning that such results would be useless to a designer. These two scenarios are illustrated in Fig. 7.

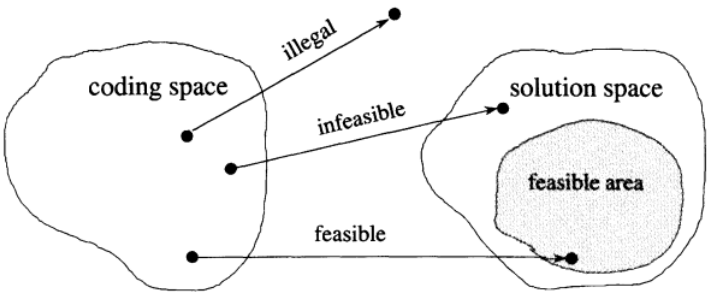


Fig. 7. Infeasible and illegal solutions in GA [30]

Penalty functions are employed to solve the problem of infeasibility. On the other hand, repair techniques are used to convert illegal solutions to legal ones [30].

GA also suffers from wrong choices of genetic operator parameters with the consequence of non-convergence and production of useless results. It is important to select an appropriate population size, rate of mutation and crossover as well as formulation of the fitness function [43].

IV. Conclusion

It is clear that over the years, both GA and precision points have overseen the synthesis four-bar for different practical application and research. A good combination of the legendary

precision point theory and evolutionary theory has shown its ability to produce very accurate results. Nevertheless, there is room for improvement into the combination to achieve even better accuracy especially with the high precessor capacities of today's computational resources. Whatever was previously viewed as impossible using the graphical and analytical methods can now be done with outmost precision. GA continues to be an important tool in modern mechanism optimisation. Since its discovery, researchers have found it handy in dimensional synthesis of different forms of the four-bar mechanism. However, its usage demands selection of certain prerequisites which cannot be ignored if the accuracy of the results obtained is to be guaranteed. Meanwhile, optimal selection of precision points is crucial in all instances. Necessary modifications to the objective functions should be considered to account for different constraints that govern the motion of the mechanisms. Furthermore, the choice of genetic operator values cannot be underrated since it also has a great effect on the accuracy of the results. Moving forward, it is high time that the combination of GA and precision points is implemented to improve real-life mechanism applications to ensure best approximations to desired output. The benefits of both techniques should be exhaustively exploited while remedies should be sought for the shortcomings that befall them. With that, the future of four-bar mechanism optimisation will be brighter than it has ever been.

From the literature reviewed, the following gaps were identified:

1) Most researchers use arbitrary number and spacing of precision points. The results obtained, thus, cannot be compared to practical mechanism designs.
2) Very few researches have been done on multi-objective four-bar mechanism optimisation.
3) Very little information is available on post-optimisation analysis of the mechanisms to evaluate their viability in practical usage.
4) Most researches only consider a section of the whole range of crank angles thus only generating a section of the coupler motion
5) The limitations of GA have not been well tackled with a view to finding remedy.
6) Most of the research is has not been implemented into practical mechanisms to evaluate the effectiveness in real-life application.

V. Abbreviations and Acronyms

CAD	Computer Aided Design
CP	Crossover Probability
DE	Differential Evolution
GA	Genetic Algorithm
GA-FL	Genetic Algorithm-Fuzzy Logic
GCP	Geometric Constraint Programming
MP	Mutation Probability
NFC	Normalised Fourier Coefficients
PSO	Particle Swarm Optimisation
TA	Transmission Angle
TE	Tracking Error

References

[1] R. L. Fox and K. C. Gupta, "Optimization Technology as Applied to Mechanism Design," *Journal of Engineering for Indsutry*, no. 72, pp. 657–663, 1973.
[2] R. Bulatovic and S. Djordjevic, "Optimal synthesis of a four-bar linkage by method of controlled deviation," *Theoretical and Applied Mechanics*, vol. 31, no. 3-4, pp. 265–280, 2004.
[3] P. Venkataraman, *Applied Optimization with MATLAB Programming*. New York: John Wiley & Sons, 1st ed., 2002.
[4] J. H. Holland, "Outline for a Logical Theory of Adaptive Systems," *Journal of the ACM*, vol. 9, no. 3, pp. 297–314, 1962.
[5] J. Cabrera, A. Simon, and M. Prado, "Optimal synthesis of mechanisms with genetic algorithms," *Mechanism and Machine Theory*, vol. 37, no. 10, pp. 1165–1177, 2002.
[6] R. Hartenberg and J. Denavit, *Kinematic Synthesis of Linkages*. New York: McGraw-Hill, 1st ed., 1964.
[7] G. P. Roston and R. H. Sturges, "Genetic algorithm synthesis of four-bar mechanisms," *Artificial Intelligence for Engineering, Design, Analysis and Manufacturing*, vol. 10, no. 05, pp. 371–407, 1996.
[8] S. S. Rattan, *Theory of Machines*. New Delhi: Tata McGraw Hill Education Private Limited, 3rd ed., 2009.
[9] L. L. Howell, *Compliant Mechanisms*. New York: John Wiley & Sons, 2001.
[10] J. S. Rao and R. V. Dukkipati, *Mechanism and Machine Theory [Textbook]*. New Delhi: New Age International Publishers, 2nd ed., 2006.
[11] F. Freudenstein, "Approximate Synthesis of Four-Bar Linkages *[1 , 2]," *Transactions of the ASME*, vol. 77, no. Aug, pp. 853–861, 1955.
[12] C. W. McLarnan, "On linkage synthesis with minimum error," *Journal of Mechanisms*, vol. 3, no. 2, pp. 101–105, 1968.
[13] I. L. Maclaine-cross, "A numerical method of Chebyshev optimum linkage design," *Journal of Mechanisms*, vol. 4, no. 1, pp. 31–41, 1969.
[14] R. S. Rose and G. N. Sandor, "Direct Analytic Synthesis of Four-Bar Function Generators With Optimal," *Journal of Engineering for Indsutry*, vol. 8, no. 72, 1973.
[15] E. C. Kinzel, J. P. Schmiedeler, and G. R. Pennock, "Kinematic Synthesis for Finitely Separated Positions Using Geometric Constraint Programming," *Journal of Mechanical Design*, vol. 128, no. 5, p. 1070, 2006.
[16] Y. Tong, D. H. Myszka, and A. P. Murray, *Four-Bar Linkage Synthesis for a Combination of Motion and Path-Point Generation*. Masters thesis, University of Dayton, 2013.
[17] H. Martínez-alfaro, "Four-bar Mechanism Synthesis for n Desired Path Points Using Simulated Annealing," in *Advances in Metaheuristics for Hard Optimization*, pp. 23–37, Berlin: Springer-Verlag Berlin Heidelberg, 1st ed., 2008.
[18] K. Mehar, S. Singh, and R. Mehar, "Optimal synthesis of four-bar mechanism for function generation with five accuracy points," *Inverse Problems in Science and Engineering*, vol. 23, no. 7, pp. 1222–1236, 2015.
[19] S. Krishnamurty, *Multiple Objective Optimization Techniques in the Synthesis and Design of Mechanisms*. Madison: University of Wisconsin, 1st ed., 1989.
[20] K. Rusell, Q. Shen, and R. S. Sodhi, *Mechanism Design:Visual and Programmable Approaches*. Boca Raton: CRC Press, 1st ed., 2010.
[21] C.-T. Ho, *Precision Position Synthesis for General Planar Mechanisms by Predictor Corrector Continuation Method*. Minnesota: University of Minnesota, 1st ed., 2001.
[22] E. Filemon, "Useful ranges of centerpoint curves for design of crank-and-rocker linkages," *Mechanism and Machine Theory*, vol. 7, no. 1, pp. 47–53, 1972.
[23] K. J. Waldron, "Elimination of the Branch Problem in Graphical Burmester Mechanism Synthesis for Four Finitely Separated Position," *ASME Journal of Engineering for Industry*, vol. 98, no. 1, pp. 176–182, 1976.
[24] K. C. Gupta and S. O. Tinubu, "Synthesis of Bimodal Function Generating Mechanisms Without Branch Defect," *Transaction of the ASME, Journal of Mechanisms, Transmissions and Automation in Design*, vol. 105, no. 82, pp. 641–647, 1983.
[25] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. Massachusetts: MIT Press, 2nd ed., 1992.
[26] M. Mitchell, *An introduction to genetic algorithms*. Massachusetts: The MIT Press, 1st ed., 1996.
[27] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*. Charlotte: Springer Science & Business Media, 3rd ed., 1996.

[28] A. Kunjur and S. Krishnamurty, "Genetic Algorithms in Mechanical Synthesis," *Journal of Applied Mechanisms and Robotics*, vol. 4, no. 2, pp. 18–24, 1997.
[29] V. M. Kureichik, S. P. Malioukov, V. V. Kureichik, A. S. Malioukov, V. I. Dempsey, and M. O. Neill, *Genetic Algorithms for Applied CAD Problems*. Berlin: Springer-Verlag Berlin Heidelberg, 1st ed., 2009.
[30] M. Gen and R. Cheng, *Genetic Algorithms & Engineering Optimization*. Ashikaga: John Wiley & Sons, 1st ed., 2000.
[31] L. Dulger, H. Erdogan, and M. Kutuk, "Matlab's GA and Optimization Toolbox: A Fourbar Mechanism Application," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 2, no. 1, pp. 10–15, 2014.
[32] M. A. Laribi, A. Mlika, L. Romdhane, and S. Zeghloul, "A combined genetic algorithm-fuzzy logic method (GA-FL) in mechanisms synthesis," *Mechanism and Machine Theory*, vol. 39, no. 7, pp. 717–735, 2004.
[33] R. Starosta, "APPLICATION OF GENETIC ALGORITHM AND FOURIER COEFFICIENTS (GA-FC) IN MECHANISM SYNTHESIS Roman Starosta," *Theoretical And Applied Mechanics*, vol. 46, no. 2, pp. 395–411, 2008.
[34] W. Y. Lin, "A GA-DE hybrid evolutionary algorithm for path synthesis of four-bar linkage," *Mechanism and Machine Theory*, vol. 45, no. 8, pp. 1096–1107, 2010.
[35] N. Nariman-Zadeh, M. Felezi, A. Jamali, and M. Ganji, "Pareto optimal synthesis of four-bar mechanisms for path generation," *Mechanism and Machine Theory*, vol. 44, no. 1, pp. 180–191, 2009.
[36] S. K. Acharyya and M. Mandal, "Performance of EAs for four-bar linkage synthesis," *Mechanism and Machine Theory*, vol. 44, no. 9, pp. 1784–1794, 2009.
[37] K. Chaudhary and H. Chaudhary, "Shape Optimization of Dynamically Balanced Planar Four-bar Mechanism," *Procedia Computer Science*, vol. 57, no. 2015, pp. 519–526, 2015.
[38] A. S. Shinde, S. A. Kulkarni, and S. S. Shete, "Dimensional Synthesis of Four Bar Mechanism Using Genetic Algorithm," *International Journal of Current Engineering and Technology*, vol. 7, no. 4, pp. 1572–1580, 2017.
[39] S. Fidanova, *Recent Advances in Computational Optimization*. Cham: Springer International Publishing Switzerland, 2013.
[40] F. Erdogdu, *Optimization in Food Engineering*. New York: Taylor & Francis Group, 2008.
[41] P. Rangaiah, *Stochastic Global Optimization: Techniques and Applications in Chemical Engineering*. London: World Scientific Publishing Co. Pte. Ltd, 1st ed., 2010.
[42] D. Reid, "Genetic algorithms in constrained optimization," *Mathematical and Computer Modelling*, vol. 23, no. 5, pp. 87–111, 1996.
[43] X.-S. Yang, *Engineering Optimization: An Introduction with Metaheuristic Applications*. Hoboken: John Wiley & Sons, 1st ed., 2010.

A Review on Marine Propeller Performance of High Speed Boat Running on an Outboard Engine

Eugene Gatete, Hiram M. Ndiritu, and Robert Kiplimo

Abstract—This paper presents a review of the marine propeller performance. Currently, water transport is very important and for that marine engines performance is necessary. The use of high-speed boats is increasing because of the population and navy security demands. Propeller blade cavitation, high fuel consumption, and high power absorption are the problems that threaten the life of the high-speed boat. Blade cavitation is caused by the disturbance of the water flow at the leading edge. The propeller blade cavitation affects the outboard propeller performance and leads to the induced vibration, erosion and excess noise of high-speed boats. There are several ways to decrease the blade cavitation, fuel and power consumption of outboard engine running the high-speed boats. Improvement of the propeller performance of an outboard engine is a possible solution for reducing the propeller blade cavitation, high fuel, and power consumption. This study reviews parameters that affect the marine propeller performance, blade cavitation development and its reduction, and improvement in propeller performance of the high-speed boats running on the outboard engine.

Keywords—High-speed boat, Marine Propeller performance, Propulsion system, Outboard engine

I. Introduction

THE conventional internal combustion outboard engine shown in Fig.1 is still the main modern method of marine propulsion and has variety of approaches to be adopted from a direct-injection two-stroke to a four-stroke cycle, with outputs from 2 to 557 horsepower (hp), and weights from a few kilograms to half a metric ton. There are various types of outboard engines in the market which vary in size, weight, output and design. The outboard engine is widely used as a power source for boat or ships due to their high performance, lower cost, simple installation, reduced fire risk, lower weight, higher speed, superior maneuverability, less draft giving better shoal water capability, and easier launching and beaching [1]. Yamaha, Honda, and Suzuki from Japan; and Mercury and Evinrude from America are the leading manufacturers [2]. The outboard engine can be divided into six subsystems base on their functions: the mechanical drive system, the swivel bracket system, water intakes, the propulsion system and the exteriors, as shown in Fig.2. The mechanical drive system includes the engine and transmission shafts to yield the kinetic energy. A swivel bracket system is the supporting and turning base and unloads the thrust from the propeller with its clamp fixed at the ships stern. The water intakes sea water enters at this point and flows through the engine for cooling purposes. The gear box and propeller make up the propulsion system. The exteriors are the covers, connections and decorative pieces. The design of the outboard propeller is the main topic in this review. Therefore, calculation and analysis of the open water propeller performance of propulsion system are needed.

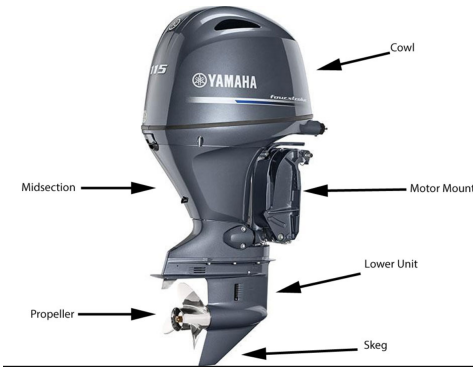


Fig. 1. Outboard Engine [2]

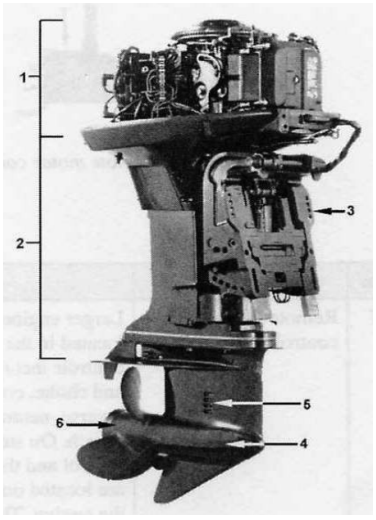


Fig. 2. Outboard Engine Parts [1]

Where in Fig. 2 ;
1 and 2: Mechanical drive system,
3: Swivel bracket system,
4: Gearbox,
5: Water intakes,
6: Propeller.

Eugene Gatete, Department of Mechanical Engineering, JKUAT (corresponding author to provide phone: +254 (0) 790249557; e-mail: eugene.gatete@students.jkuat.ac.ke).

Hiram M. Ndiritu, Department of Mechanical Engineering, JKUAT (e-mail: hndiritu@eng.jkuat.ac.ke).

Robert Kiplimo, Department of Marine and Maritime Operations, JKUAT (e-mail: kiprob@eng.jkuat.ac.ke).

II. Outboard Propulsion System

The propulsion system of a boat is an assembly of components to drive boats based on rotary motion. One of the most critical propulsion systems is the propeller, which produces the power needed to push water while the main engine is running, as shown in Fig. 3. Its configuration involves a mechanism to push water, with the resultant reaction propelling the boat forward. It plays a great role in the fuel efficiency, thrust, torque and overall efficiency improvement. Due to its role in the performance of the boat as shown in Fig. 4, it is necessary to predict the performance of the considered propeller.

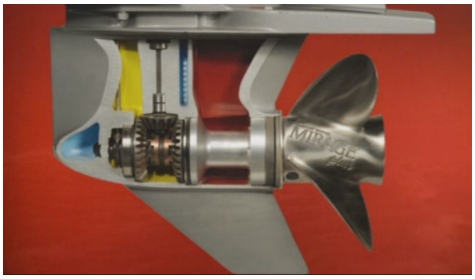


Fig. 3. Propulsion system of an outboard engine [13]

Marine propellers face the problems of blade cavitation which leads to erosion, noise, and vibration. Meanwhile, the propeller configuration influences fuel consumption, efficiency as well as performance of high speed boats [13].

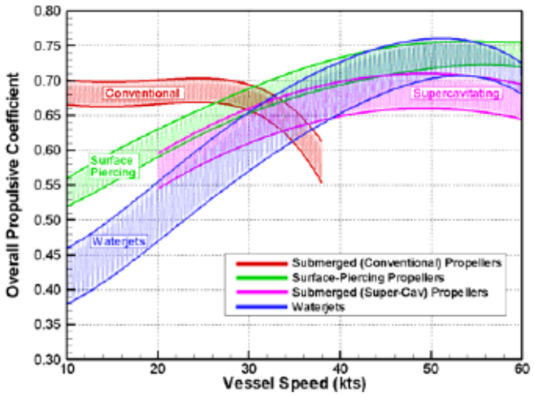


Fig. 4. Overall propulsive coefficient versus ship speed for different propulsor types [12]

A. Outboard Propeller

Propeller design is used to obtain a blade geometry which meets the requirements determined by the operating condition by using various design methods [7], [8]. The optimization of the outboard propeller is to provide the maximum thrust for the minimum torque at a specific rotational velocity (rpm) with a particular boat speed. Initially, propeller design intended to achieve the highest efficiency. Then, the cavitation phenomenon came into focus and now propeller design involves the two factors; efficiency, and cavitation.

Choosing the right propeller is crucial in determining the performance of the outboard engine. Propeller choice can affect boat top speed by as much as 5 to 10 knots. It also has a direct effect on acceleration, cornering, pulling power and fuel economy. Some boats may require change of propellers for different activities, such as high speed cruising, water skiing, or carrying heavy loads. Using the wrong propeller in any of these applications will not only hurt performance, but could also cause engine damage [14]. There are also many types of propeller systems in operation. The right hand fixed blade propeller is still common. Meanwhile, development of controllable pitch propellers, contra-rotating propellers, multi-blade propeller systems, twin, triple and quadruple propeller sets, pod propulsion units, Kort nozzle systems and Azipod systems have also all taken market share in both commercial and boat construction [3].

The propeller is a powerful propulsion device used in a boat. Furthermore, utilizing propeller in high speed boats running on an outboard engine avoids any other damages. Its performance is conventionally represented in terms of non-dimensional coefficients like the thrust coefficient, torque coefficient, and efficiency and their variations with the advance coefficient [15].

It is difficult to determine the characteristics of a full-size propeller in open water by varying the speed of the advance and the revolution rate over a range while measuring the thrust and torque of the propeller, but with the advent of computers, numerical methods developed rapidly since the 1960s onwards. The first numerical methods were based on the lifting line theory and later the lifting surface model were developed by Salvatore et al [16]. Later on, a propeller performance analysis program was also developed and integrated into a genetic algorithm by Burger et al [17].

Sanchez et al. [18] calculated open water flow patterns and performance coefficients for DTRC 4119 propeller using FINFLO code. The flow patterns were predicted by the k-turbulent model. The authors also suggested a better prediction of the tip vortex flow, which required a more sophisticated turbulence model.

Motley et al. [19] investigated a reliability-based global design of self-adaptive marine propellers operating under a range of steady loading conditions, using a Nelder-Mead constraint based optimization technique. An optimized propeller of Vo= 13.0 knots and eq = 17:250 was found to reduce the load variation by approximately 10%, cavitation potential by 2.3% on the back side and 9.7% on the face side. It also increased the total efficiency by approximately 0.3%. For unsteady applications, the authors suggested improvements on propeller performance in terms of reduction of loads, structural strength, delay of cavitation and hydrodynamic efficiency.

B. Geometrical Parts of Outboard Propeller

Marine propeller is a set of identical twisted blades, spaced evenly around a hub. Most propellers have a splined bushing in the hub that mounts on the outboard. Fig. 5 bellow shows an example of marine propeller used in outboard engine [45].

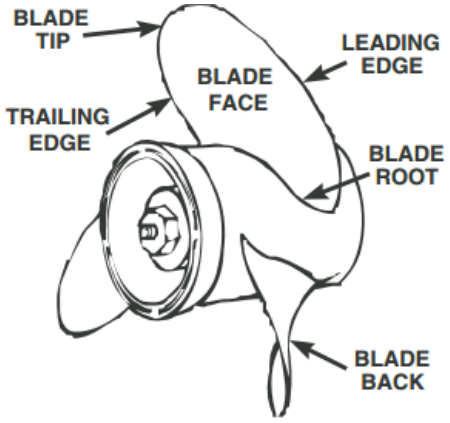


Fig. 5. Parts of an outboard propeller [45]

1) **Hub:** The boss of a propeller is the solid center disc, to which the propeller blades are attached. Since the hub generates no drive, the ideal would be to eliminate it. As a practical matter, though, the hub can seldom be much less than percent of the diameter in order for it to have sufficient strength.
2) **Blades:** The propeller blades are the twisted fins aor foils that project out from the hub. It is the action of the blades that drives a boats through the water.
3) **Blade Face and Blade Back:** The blade face is the high-pressure side, or pressure face of the blade. It is the side facing aft, the side that pushes the water when the boat is moving forward. The blade back is the low pressure side or suction face of the blade, the side facing ahead.
4) **Blade Root and Blade Tip:** The blade root is the point at which the blade attached to the hub. The blade tip is the extreme outermost edge of the blade, as for from the propeller shaft center as possible.
5) **Leading and Trailing Edges** The leading edge of a blade is the edge of the blade that cleaves the water. The trailing edge is the edge from which the water streams away.

III. High-Speed Crafts

A high-speed craft (HSC) shown in Fig. 6 is a vessel for civilian use, also called a fast craft or fast ferry and is called patrol craft for military purposes. A vast increase in the high-speed crafts due to existing needs in the field of fast transport of light and expensive cargo, passengers at the high-speed craft for marine transportation has drawn considerable interest for both ship owners and naval architects. The function of high speed also gives advantages to boats which are designed to be used for a surveillance and patrol in maritime area at open sea. The advanced concept was applied in many types of high speed craft in order to obtain a great performance in Seaway. The design and safety of high-speed craft is regulated by the high Speed Craft Codes of 1994 and 2000, adopted by the Maritime Safety Committee of the International Maritime Organization (IMO) [4].

Savitsky et al. [5] defined that high speed craft are considered to be vessels that can travel at a sustained speed equal to or greater than 35 knots with bursts of high speeds of 40-60 knots. Froude number allows for another way to hydrodynamically classify ships. Naval architects use the Froude number when vessels deal with the interaction of the waters free surface and the hull. High speed vessels are typically defined with Froude number greater than 0.4 which at this speed range the crafts weight is almost entirely supported by dynamic forces and the trim tend to be much lower than hump. It is classified into two categories which are an Air-Supported and Displacement type. Air supported crafts include Air Cushion Vehicles (ACV), Surface-Effect Ships (SES) and Foil Supported craft such as hydrofoils and jetfoils. Displacement type vessels include conventional monohull, catamaran, trimaran, small water plane area twin hull (SWATH), and air lubricated hulls. Each type of craft has its unique characteristics, and they all suffer from the common problem of limited payload and sensitivity to wind and sea state. Besides, they are vessels with a design speed corresponding to a Froude number above 0.45, primarily designed for short distance services such as public transport of passengers and vehicles. An outboard engine, inboard diesel sterndrive, or inboard diesel waterjet can power the high-speed watercraft [9], [10]. Therefore, the improved performance of the high speed boat for the local market is to enhance the operations in green or blue water for support in short distance operations, border actions, and international peacekeeping or disaster relief [11].

Fig. 6. Example of high speed craft (HSC) [11]

Black, et al. [12] developed new blade section concepts that have the efficiency characteristics of conventional submerged sub-cavitating propellers at low and intermediate speeds but can transition to a super-cavitating mode for high speed operation without encountering thrust breakdown.

IV. Marine Propeller Performance Characteristics

A propeller is normally fitted onto the lower unit of an outboard engine where it runs in water that has been disturbed by the boat as it moves ahead. The performance of the propeller is thus affected by the boat to which it is fitted. Hence, in order to determine the performance characteristics of a propeller unaffected by the boat to which it is fitted, it is necessary to make the propeller operate in open water. Therefore, some of the parameters that can affect the performance while designing are subdivided into two

categories, that is, open water characteristics and propeller-hull interaction characterization [20].

A. Open Water Characteristic

The forces and moments produced by the propeller are expressed based on non dimensional characteristics. These non-dimensional terms explaining the general performance characteristics are established using dimensional analysis [20]. Thrust (T) and Torque (Q) can be represented by the following functions depending upon the physical quantities involved;

$$T = f_1(\rho, D, V, g, N, P, \mu) \tag{1}$$

and

$$Q = f_2(\rho, D, V, g, N, P, \mu). \tag{2}$$

An efficient propulsion system relies on propeller performance which depends on the thrust force, torque, and efficiency. These were determined by non-dimensional quantities through propeller performance, which are, thrust coefficient(Kt), torque coefficient (Kq), coefficient of cavitation and open propeller efficiency(η_o) with respect to the advance coefficient (J). The expressions for K_T,K_Q, σ, η_o and J are given in Equations 3 to 8 as shown by Techet et al [21], [22].

$$K_T = \frac{T}{\rho N^2 D^4} \tag{3}$$

$$K_Q = \frac{Q}{\rho N^2 D^5} \tag{4}$$

$$\sigma = \frac{P_0 - P_v}{2\rho N^2 R^2} \tag{5}$$

$$J = \frac{V_a}{ND} \tag{6}$$

The efficiency of the propeller is the ratio of useful power produced by the propeller, the thrust horsepower (THP), to the input shaft power, the delivered horsepower (DHP):

$$\eta_o = \frac{THP}{DHP} = \frac{TV_a}{2\pi NQ} \tag{7}$$

Additionally, the efficiency can be written as:

$$\eta_o = \frac{K_T \rho N^2 D^4 V_a}{2\pi N K_Q \rho N^2 D^5} = \frac{K_T}{K_Q} \cdot \frac{V_a}{2\pi ND} = \frac{K_T}{K_Q} \cdot \frac{J}{2\pi} \tag{8}$$

Where;
T= Thrust [N],
Q= Torque [Nm],
D= Diameter [m],
V_a=$J.N.D$ (Advanced Speed)[m/s],
N= Rate of rotation[rps],
ρ= Mass density of water[kg/m^3],
μ= Dynamics viscosity of water[Pa/s],
P= Total static pressure[Pa],
P_o= Absolute pressure at shaft centre [Pa],
P_v= Vapour pressure at ambient temperature [Pa].

Taheri et al. [23] developed a propeller design method based on a vortex lattice algorithm to optimize the shape and efficiency. The analysis of the hydrodynamic performance parameters based on a vortex lattice method was used to implement two computer code. The first code was sequential unconstrained minimization techniques for minimizing the torque coefficient considering the thrust coefficient constant as a constraint and chord distribution as a design variable. The second was a modified genetic algorithm to maximize efficiency by considering the design variables as non-dimensional blades chord and thickness distribution along the blade. The comparison between experimental data and the solution of the optimization problems were found to be near 13% improvement in efficiency and a nearly 15% decrease in torque coefficient for the propeller.

Senthil et al. [24] investigated a computational method for the determination of open water performance of a marine propeller using the RANS solver and unstructured meshes and compared the results to experimental ones. The result of thrust, torque and open water efficiency found using CFD (based on RANSE) were not accurate compared to the existing propeller chart. This inaccuracy was due to the challenge of handling a large number of cells in the flow domain and incorporating a rotating flow over propeller simulation. The author suggested the structured meshes and proper modeling of the hub to get an accuracy result of thrust, torque and open water efficiency in Computational Fluid Dynamics.

Shiu-Wu Chau et al. [25] investigated the hydrodynamic performance of high-speed craft rudders via turbulent flow computations with non-cavitating characteristics, by focusing on the numerical analysis of non-cavitating hydrodynamic characteristics of practical rudders used for high-speed crafts, which is valid in the low-speed region. Thrust, torque, and efficiency, the lift, drag and stock moment coefficient of rudder were evaluated to investigate the influences of profile shape and profile thickness. The result showed that, the location of maximum thickness is the most important factor to influence the non-cavitating hydrodynamic characteristics.

B. Propeller Hull Interaction

Kim et al. [26] in their studies on high-speed planning hulls for improving the seaworthy performance, designed three planning hulls namely deep-V planing straight (VPS), deep-V wave-piercing concave (VWC), and deep-V wave-piercing straight (VWS) having almost the same displacement and principal dimension. The hydrodynamic characteristics were compared with each other through the model tests. It was found that among the three model boats, VWS model had favorable seaworthy performance due to its reduced wetted surface area. However, its hull form needed to be optimized in order to improve its hydrodynamic performance.

An experimental and numerical investigation on hydrodynamic and aerodynamic characteristics of a plan boat was performed by Jiang et al. [27]. The comprehensive series of viscous CFD simulations considering free-surface and 2-DOF motion of the hull was used. The result showed that the variation between the calculated and experimental resistance increased from 2.99% to 14.66% as the Froude number increased from 3.16 to 5.87. In the numerical simulation, the total resistance increased up to the value of

Froude number for which wave surface separated from the tunnel roof and the total resistance decreased. Finally, the calculated results were validated by comparing experimental and numerical data and this showed a good agreement.

V. Parameter Affecting the Outboard Propeller Performance

There are three most significant factors affecting the propeller performance and efficiency which include diameter, rotational speed, and pitch. Pitch is the distance that a boat is propelled forward in one propeller rotation. Many other variables are needed for selection of suitable propeller which include the pitch to diameter ratio, rotation speed, blade number, blade area ratio, skew angle, blade shape and blade thickness [1], [6].

A. Propeller Diameter

The diameter is a crucial geometric parameter in determining the amount of power that a propeller can absorb and deliver, and thus dictating the amount of thrust available for propulsion. With the exception of high speed (35 Knots+) vehicles the diameter is proportional to propeller efficiency (i,e. Higher diameter equates to higher efficiency). In high speed vessels, however, larger diameter equates to high drag. For typical vessels a small increase in diameter translates into a dramatic increase in thrust and torque load on the engine shaft, thus the larger the diameter the slower the propeller will turn, limited by structural loading and engine rating. For optimum efficiency, the propeller diameter determination is a complex procedure involving various empirical formulas [6]. In high speed craft, with regards to cavitation, a larger diameter is beneficial as the rotational speed of the propeller could be reduced and still achieve the required forward movement. A reduced rotational speed would mean that the pressure unbalance on the blades would be reduced with the reduction of the inflow forces and so cavitation would be decreased. So for a certain engine output or desired forward speed, a large diameter propeller would allow for slower rotation and reduced cavitation [3].

B. Rotational Speed

For vessels operating under 35 Knots speed, reduction of speed and increase in diameter results to higher torque. For high speed boat, RPM is ranging from 2000 to 6000 rpm and from that it is possible to get the inflow velocity for a given advance coefficient [28].

C. Pitch to Diameter Ratio

The pitch of a propeller indicates the distance the propeller would drive forward for each full rotation. In reality since the propeller is attached to a shaft it will not actually move forward, but instead propel the ship forward. The distance the boat is propelled forward in one propeller rotation is actually less than the pitch. For high speed craft, pitch ranges from 9 inch to 24 inches. Typically blades are twisted to guarantee constant pitch along the blades from root to tip. Often a pitch ratio will be supplied. This is simply the ratio of pitch to diameter, usually in millimeters, and typically falls between 0.5 and 2.5 with an optimal value for most vessels closer to 0.8 to 1.8. Pitch effectively converts torque of the propeller shaft to thrust by deflecting or accelerating the water astern simple Newtons Second Law [6], [29].

Mojtaba et al. [30] designed a marine propeller to generate the thrust with lower torque, highest efficiency and reducing cavitation. The author used numerical method based on Blade Element Theory (BET) to get the thrust and analyze how it depends on the shape of the marine propeller. The parameters such as pitch ratio, blade area ratio and skew angle were used as the input variables to achieve the optimum propeller with high propeller performance. It was found that various pitch ratios such as 0.4, 0.6, 0.8, 1.0, 1.2, and 1.4 used satisfy the propeller performance characteristics. The skew angle effect also showed its capacity to estimate the propeller efficiency in limitation of cavitation problem. This research did not consider diameter, blade number and material selection for optimization.

Kiam et al. [31] investigated the marine propeller performance characterization through CFD to predict a three blade marine propeller performance characteristics. Five propellers with a pitch to diameter ratio values of 0.6; 0.8; 1.0; 1.2; and 1.4 were used for the computational flow analysis through RANS solver to compute the propeller performance characteristic to the advance coefficient(J). It was found that efficiency, torque, and thrust increased as the pitch diameter ratio increased.

M.Bernitsas et al. [32] designed the Wageningen B-Series propeller to test the open water characteristics using the multiple polynomial regression analysis. The derived polynomials expressed thrust and torque coefficients in terms of the number of blades, the blade area ratio, the pitch-diameter ratio and the advance coefficient. It was shown that the derived polynomials valid for the pitch-diameter ratio varying between 0.5 and 1.4 for the Reynolds number of 2×10^6. It was also found that the thrust coefficient displayed a local maximum for high pitch-diameter ratio, high number of blades, low blade area ratio and low values of advance coefficient. However, the extremes of the above ranges were not considered in this research.

D. Number of Blades

Blades are the twisted fins or foils that protrude from the propeller hub. The shape of the blades and the speed at which they are driven dictates the torque a given propeller can deliver. The primary effect that is to be avoided with propeller blade number selection is resonance, as the number of blades affects the frequency of vibrations and the strength of vibrations that occur during operation. There is also a strong interrelation between propeller diameter, blade area and blade number, where a greater diameter requiring fewer blades and a greater area requiring more blades. An increase in propeller number can reduce sheet cavitation of the suction side due to a reduced load per blade, however it can increase root cavitation due to reduced clearance between each blade [29]. It is recommended

that the number of blade should be in range of 2 to 6 blades to reduce the resonance. An area ratio of 16-18% should be also assigned to each blade, with a decrease of 4% in the diameter per each additional blade. This guideline has been obtained with both performance and cavitation efficiency in mind [33].

Kiam et al. [34] also investigated the effects of number of blade on marine propeller performance. Five propellers with number of blades of 2, 3, 4, and 5 were utilized for the computational flow analysis through RANS solver to investigate the propeller performance. It was seen that the efficiency decreased as number of blades increased while torque and thrust increased as the number of blades increased. It was concluded that, although higher number of blades reduced the efficiency and leads to higher fuel consumption but it provided better velocity.

Boucetta et al. [35] carried out a numerical simulation of the flow around marine propeller using RANS method of the commercial CFD code fluent to examine the effect of skew magnitude, thickness and blade number on the hydrodynamic performances. The number of the blades were changed from three to four and five to investigate its effect on the open water characteristics of the marine propeller. It was concluded that the increase of blade thickness generated a rise in propeller efficiency, and the adoption of skew angle on the blade improved the hydrodynamic performance of marine propeller. In addition, the propeller with four blades gave the best efficiency. However, this work did not consider the effect of blade materials, pitch diameter ratio and rotational speed on propeller performance.

Xueyin Wu [28] in his development process for marine propeller through design, simulation and prototyping explained the effect of the number of blade on propeller performance. The author showed that propellers normally have a number of blades within the range from two to six. In his study, it was found that more number of blades increased thrust which caused cavitation and which also led to propeller vibration. It was also shown that less number of blades avoided cavitation and decreased thrust. Therefore, the author concluded that, a good propeller has a large diameter, slow speed, low number of blades and high efficiency.

VI. Marine Propeller Blade Cavitation on Propeller Performance

Propeller cavitation is phenomenon which begins when a disturbance creates a low-pressure area in the water flow. As speed increases, the low pressure intensifies enough to vaporize (boil) some of the surrounding water. When the vapor bubbles approach a high pressure area, they collapse, releasing energy and causing damage. Cavitation is caused by a disturbance of the water flow in front of the propeller. It is also caused by an irregularity in the boat bottom or gearcase, and a misplaced transducer or speedometer pickup [36]. The results of cavitation usually appear as burned areas on the propeller blades as it is shown in Fig. 8;

Propeller Cavitation shown in Fig.7, also referred to as ”cold-boiling”, is the phenomenon of the formation of vapour

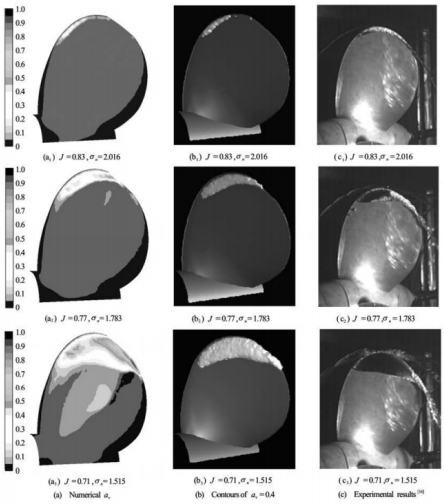


Fig. 8. Variation of propeller blade cavitation [47]

pockets -cavities- within a fluid, caused by pressure reduction below a certain value named vapour pressure”. When the cavities find themselves back in a higher pressure environment they implode, causing the liquid to rush towards its center to fill it, hence generating large pressures (up to 1 GPa). These are subsequently propagated in the form of pressure waves to their surroundings [48].

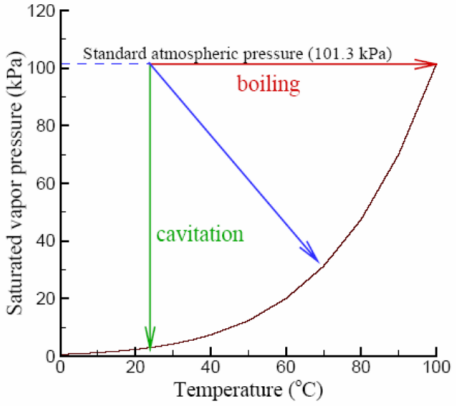


Fig. 7. Difference between cavitation and boiling phenomena [48]

In marine hydrodynamics this phenomenon is especially present in pumps, turbines and propellers, which often comprise of multiple hydrofoils. As pressure is a decreasing function of velocity, according to Bernoulli, the flow around each propeller blade section experiences rapid pressure changes, which lead to the inception of cavities around them. This effect is amplified at high inflow speeds but is also dependent on the body geometry [2], [40].

This phenomenon occurs also in many fields such as automotive industry, chemical engineering, cleaning technologies, the biomedical applications and, of course, has developed a dedicated research branch under fluid mechanics. The effect of cavitation on propellers was first investigated by Reynolds et al. [39] in the laboratory, and by Barnaby et al. [41] using full scale trials of the destroyer daring. They found

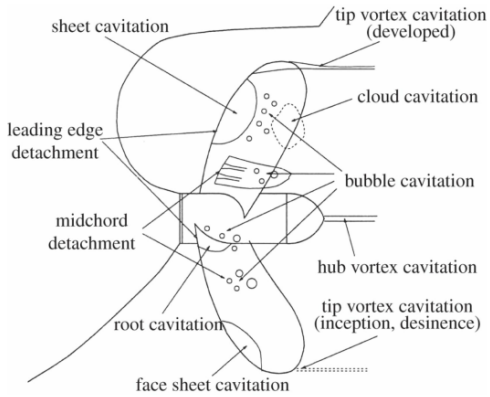


Fig. 9. Various types of cavitation [44]

that the formation of vapor bubbles on the blades reduced the power of the propeller. Later investigators also found that cavitation can lead to undesirable effects such as blade surface erosion, increased hull pressure fluctuations and vibrations, acoustic energy radiation, and blade vibration.

Propeller cavitation has been found to manifest itself in different locations of the propeller and with certain unique forms as show in Fig. 9. Thin string-like cavities at the tip of the blade and from the hub of the propeller are called vortex cavities and are fully developed. On the other hand, the sheet cavities encounter were relatively stable and present the least modeling difficulties in numerical methods. Sheet cavities start from the leading edge of the blade whereas if they start from mid-chord they tend to turn into bubble cavitation from there and aft ward. Root cavitation may appear just forward of the hub at certain rotation angles of the propeller and often joins the hub vortex. Finally, cloud cavitation is visually similar to sheet cavitation, however is very unstable and presents many modeling difficulties. Other type of cavitation which can also occur include bubble cavitation [40].

In Yin Lu Young's study [44], the cavity on a propeller blade is treated strictly as sheet cavitation. She states that the pressure inside the sheet cavity is assumed to be constant and equal to the vapor pressure. The rationale behind using the sheet cavity model includes: It provides a relatively simple mathematical model where potential flow theory can be applied. Tulin et al. [43] found that sheet cavity is the first-order contributor to dynamically varying blade loads; and other forms of cavitation (such as tip or hub vortex cavitation) and other neglected phenomena (such as wake roll-up) can be added as refinements to the current models.

Kuiper et al. [42] proposed propeller design techniques to delay the tip vortex cavitation inception. Important parameters in the tip region including thickness, plan form, skew, chord distribution and rake were systematically varied while maintaining a constant radial loading distribution. A systematic series of 2 bladed propeller designs was evaluated using a panel method for pressure distribution near the tip. An extreme tip rake towards the pressure side was used in the investigation. The measurements showed a trailing vortex coming from the corner of the raked tip. There was still too much cross-flow over the area of strong curvature, leading to separation and vortex formation. Local and leading edge tip vortex inception were delayed significantly, while the width of cavitation bucket for the trailing tip vortex inception was reduced moderately. The authors believe that the strong curvature in the tip region should be avoided.

Gaggero et al. [46] designed a propeller for a high-speed craft using a multi-objective numerical optimization approach based on boundary element method. The efficiency was numerically investigated using RANSE calculation. The design improved the propulsive efficiency, reduced blade cavitation and simultaneously maximized the ship speed. It was also found that the proposed approach was able to deal with main objectives and to derive trade-off designs with high performance simultaneously concerning suction and pressure side cavitation. However, the author did not investigate the improvement of the quality of the design regarding the risk of leading edge, the pressure side, and blade cavitation.

Numerically, Zhi-feng et al. [47] investigated the cavitation and hydrodynamics performance of the propellers to predict the thrust, the torque and the vapour volume fraction on the back side of propeller blade for a uniform inflow. The authors used a viscous multiphase flow theories based on RANSE formulation of advance rate (J) and cavitation number (n) to validate the result. It was found that for the high value of advance rate the cavitation was relatively weak and had the little effect on the hydrodynamic performance while the sixteen small value of advance rate the cavitation was strong. In addition, the increase of cavitation number improved the propeller performance characteristics. However, Blade number and area ratio, rotational speed, and blade material were not considered in this work.

VII. Conclusion

1) From the review, it is evident that sufficient research has not been done to determine the optimum propeller design for high-speed boats running on an outboard engine while reducing the propeller blade cavitation as low as possible. Hence, there is a need to improve the quality of propeller design in the cases of the original rake distribution with regards to the risk of leading edge, the pressure side, and blade cavitation.
2) Due to limited literature on propeller geometry and material, there is a need to investigate the effect of propeller geometry such as pitch diameter ratio, rotational speed (rpm), and blade number on propeller performance.
3) The use of outboard engine as an alternative propulsion for high speed boat remains a challenge due to the propeller blade cavitation. Thus, an investigation on the parameter affecting the propeller blade cavitation is needed to get a marine propeller with the high-performance application.

References

[1] H. S. Rainbow, "Some notes on outboard motors," Inter J Nav Arch. Oc Eng., vol. 178, no. 1, 1963

[2] Sterling, Frank Ward, ed. *"Marine Engineers Handbook."* McGraw-Hill Book Company, Incorporated, 1920
[3] Johnson and Evinrude,*"Propeller Selection Guide"*, researchgate.net, 2011
[4] The Maritime Safety Committee, *"International Code of Safety for High-Speed Craft"*, International Maritime Organization, 1994
[5] Savitsky D, and Brown, P.W *"Procedure of Hydrodynamic Evaluation of Planning Hull in smooth and rough water"*, Marine Technology, 13(4), 381-400 Oct 1976
[6] MAN Diesel Turbo, *"Basic Principles of Ship Propulsion,"* Marinelink, 2011.
[7] Carlton, John. *"Marine propellers and propulsion."* Butterworth-Heinemann, 2012.
[8] Kerwin, Justin E. *"Hydrofoils and propellers."* Lecture Notes, Department of Ocean Engineering, Massachusettes Institute Technology, USA (2013)
[9] ITTC, *"Recommended Procedures and Guidelines of ITTC,"* ITTC media, pp. 18, 2011
[10] R. Yousefi, R. Shafaghat, and M. Shakeri, *"Hydrodynamic analysis techniques for high-speed planing hulls,"* Elsevier, vol. 42, pp. 105(113), 2013
[11] J. Haynes, *"Innovative Propulsion Systems for Fast Craft,"* marinelink, 2016
[12] Black, S., Shen, Y. and Jessup, D. *"Advanced Blade Sections for High Speed Propellers",* Proc., Propeller/Shafting 2006 Symp., SNAME.
[13] T. K. Yoshitaka Nishihara, *"Development of Estimation Method of Fuel Consumption Performance in Lifetime for High-speed Craft with Outboard Engine,"* Researchgate, no. July, 2010.
[14] C. Barry, *"Propeller Selection For Boats and Small Ships,"* Engines, Mar., pp. 1(32), 2005.
[15] H. S.Subhas, V.F.Saji, S.Ramakrishna, *"CFD Analysis of a Propeller Flow and Cavitation,"* Int. J. Comput. Appl., vol. 55, no. 16, pp. 2633, 2012
[16] S. Ianniello, *"Theoretical Modelling of Unsteady Cavitation and Induced Noise,"* Italy. Natural. Res. Counc., vol. 37, 2015.
[17] C. Burger, John E. Burkhalter, Roy J. Hartfield, Robert S. Gross, and Ronald M. Barrett, *"Propeller performance analysis and multidisciplinary optimization using a genetic algorithm,"* Auburn University, Alabama, vol. Aerospace, no. Ph.D. Diss., 2007
[18] A. Sanchez-caja, *"DTRC Propeller 4119 Calculations at VTT,"* VTT Tech. Res. Cent. Finl., vol. 30, 2015.
[19] Y. L. Michael R.Motley, *"A reliability-based design and optimization of self-twisting composite marine propeller,"* Eur. Fluids Eng. Summer Meet., 2017
[20] K. J. Rawson and E. C. Tupper, *"Basic Ship Theory,"* Researchgate, vol. 2, 2010
[21] A.H.Techet,*"Propeller Notes"*, Hydrodyn. Ocean Eng., p-1-20, 2004
[22] John, Carlton,*"Hydrodynamic Characteristics of Propellers"*, Int. Shipbuild. Prog., p.1-16, 2012
[23] R.Taheri, K.Mazaheri, *"Hydrodynamic optimization of marine propeller using gradient and non-gradient based algorithms,"* International Journal of Advance Research in Science and Engineering, vol. 3, 2013
[24] M.N.Senthil Prakash and R.Deepthi Nath, *"A computational method for determination of open water performance of a marine propeller,"* Int. J. Comput. Appl., vol. 58, no. 12, pp. 15, 2012
[25] Chau, Shiu-wu and Kouh, Jen-shiang and Wong, Teck-hou and Chen, Yen-jen,*"Investigation of Hydrodynamic Performance of High-Speed Craft Rudders via Turbulent Flow Computations"*,J. Mar. Sci. Technol.,1 pp.61-72, vol.13, 2005
[26] D. J. Kim, S. Y. Kim, Y. J. You, K. P. Rhee, S. H. Kim, and Y. G. Kim, *"Design of high-speed planing hulls for the improvement of resistance and seakeeping performance,"* Inter J Nav Arch. Oc Eng, vol. 5, no. 1, pp. 161-177, 2013
[27] Y. Jiang, H. Sun, J. Zou, A. Hu, and J. Yang, *"Experimental and numerical investigations on hydrodynamic and aerodynamic characteristics of the tunnel of planing trimaran,"* Appl. Ocean Res., vol. 63, pp. 1-10, 2017.
[28] X. Wu, *"A Rapid Development Process for Marine Propellers through Design, Simulation and Prototypimg,"* Sci. Appl., 2010
[29] C. John, *"Marine Propellers and Propulsion.,"* Oxford: ButterworthHeinemann, vol. 2, 2012
[30] M. Kamarlouei, H. Ghassemi, K. Aslansefat, and D. Nematy, *"Multi-Objective Evolutionary Optimization Technique Applied to Propeller Design,"* Acta Polytech. Hungarica, vol. 11, no. 9, pp. 163182, 2014
[31] K. B. Y. Wai Heng Choong and W. Y. Hau, *"Wageningen-B Marine Propeller Performance Characterization Through CFD,"* Appl. Sci. 14, vol. 11, 2014.
[32] D. M.M Bernitsas, *"Kt, Kq and Efficiency Curves for the Wageningen B-Series Propellers,"* Agris, vol. 237, p. 115, 1981.
[33] E. Johnson, *"Propeller Selection Guide,"* Johnson, Evinrude Genuin. Parts, 1998
[34] C. M. O. Kiam Beng Yeo, Rosalam Sabatly and W. Y. Hau, *"Effects of Marine Propeller Performance and Parameters Using CFD Method,"* J. Appl. Sci., vol. 22, 2014
[35] D. Boucetta and O. Imine, *"Numerical Simulation of the Flow around Marine Propeller Series,"* J. Phys. Sci. Appl., vol. 6, no. 3, pp. 55-61, 2016
[36] RINA, *"Cavitation of Propellers,"* Stone Mar., 2011
[37] B. S. Massey and A. J. Ward-Smith, *"Mechanics of Fluids."* London; New York: Taylor Francis, 2006
[38] J. N. Newman, *"Marine Hydrodynamics."* MIT Press, 1977.
[39] Reynolds, O. *"'On the effect of immersion on screw propellers."* Transactions of Institute of Naval Architecture, 2, 1874.
[40] Kosal, E. *"Improvements and enhancements in the numerical analysis and design of cavitating propeller blades."* Masters thesis, UT Austin, Dept. of Civil Engineering. Also, UT Ocean Eng. Report 99-1, 1999.
[41] Barnaby, S. W.*"On the formation of cavities in water by screw propellers at high-speed."* Transactions of Institute of Naval Architecture, 38, 1897.
[42] S.-W. Chau, K.-L. Hsu, J.-S. Kouh, and Y.-J. Chen, *"Investigation of Cavitation Inception Characteristics of Hydrofoil Sections via a Viscous Approach",* Journal of Marine Science and Technology, vol. 8, no. 4, pp. 147-158, 2004
[43] Tulin, M. *"An analysis of unsteady sheet cavitation."* In The 19th ATTC Conference, pages pp. 10491079, 1980.
[44] Kuiper, G., van Terwisga, T.J.C., Zondervan, G.J., Jessup, S.D. and Krikke, E.M., *"Cavitation inception tests on a systematic series of two blade propellers",* Rome, Italy, Proc., 26th Symp. on Naval Hydrodynamics, 2006.
[45] D. R. Smith and J. E. Slater, *"The Geometry of Marine Propellers",* Defence Research Establishment Atlantic, Dartmouth, Nova Scotia, 1988
[46] S. Gaggero, G. Tani, D. Villa, M. Viviani, P. Ausonio, P. Travi, G. Bizzarri, and F. Serra,*"Efficient and multi-objective cavitating propeller optimization : An application to a high-speed craft,"* Phys. Procedia, vol. 64, pp. 3157, 2017.
[47] Z.-f. Zhu and S.-l. Fang, *"Numerical Investigation of Cavitation Performance of Ship Propeller,"* J. Hydrodyn., vol. 24, no. 3, pp. 347353, 2012.
[48] F. Salvatore, L. Greco, and D. Calcagni, *"Computational analysis of marine propeller performance and cavitation by using an inviscid-flow BEM model,"* no. June, 2011.

A Review of Ball mill grinding process modeling using Discrete Element Method

Philbert Muhayimana, James K Kimotho, and Hiram M Ndiritu

Abstract—In the two past decades, the discrete element method (DEM) has been used to model the grinding mill and grinding media motion in mineral processing industries. There have been many different discrete element method models that provided good prediction of granular material behavior mainly the trajectory of particles, contact force, kinetic energy, and power draw of ball mill. This paper is intended to review the ball milling parameters that affect the grinding performance of a ball mill mainly the power consumption and the throughput quality. Particularly, the discrete element method (DEM) and its use to optimize critical parameters, limitations and achievement and identifying areas that still need further research.

Keywords—Ball mill, comminution, discrete element method, tumbling mill

I. INTRODUCTION

IN mineral processing, valuable ore minerals need to be liberated from the gangue in order to achieve a product with desirable grade after concentration processes. The release of these valuable minerals is obtained through comminution which is done in a grinding mill or a crusher machine. The process of size reduction also known as comminution is highly energy intensive and expensive. It is estimated that industrial comminution processes absorb from 3 to 5% of global electric energy consumption [1]. Fig. 1 shows the typical cost of grinding mill influenced by energy consumed, type of liner used, and the grinding media. The total milling cost (energy, grinding media, liner/lifters, and labor cost) is affected by the mill liner/lifter particularly in autogenous (AG) and semi-autogenous(SAG) the energy usage in mining industries [2]. The total grinding cost was found to be 40 of the total cost in mining industries. Ball mill grinds material by rotating a cylinder with grinding balls, causing the balls to fall back into the cylinder and onto the material to be ground as shown in Fig. 2 [3]. Collision impact reduce particle size as a result of kinetic energy and potential energy of balls. Grinding mills are able to reduce size particles on a relatively wide range of particle sizes, hence, their wide applicability in the industry, production of noncrystalline materials and in research laboratories [4]. The movement of grinding media and granular material is affected by different parameters: the design of the mill drum (the mill diameter, the length, the size of the liner, and lifters); grinding media (ball size and their size distribution, ball material); Mill filling (the charge volume, feed size) critical speed and grinding time will affect the efficiency of the ball mill [5]. Researches aimed at understanding the grinding parameters that affect the grinding mechanisms have been conducted, estimating power draft, and modeling milling process. Discrete element method (DEM) has made a great contribution in modeling and understanding the grinding problem. For instance, DEM can be used to model the collisions of individual materials in the drum, which when applied to the entire charge mass over a period of time results in the mass charge motion. DEM is also quite reliable because the underlying principles originate from the fundamental laws of physics, provides an insight into the charge motion, and simultaneously gives other information, such as distribution of impact energy, force transmission, and stresses on the wall, etc [6].This paper seeks to review different methods used to model the grinding process of a ball mill as well as to evaluate the effect of lifter geometry, grinding media size and mill speed on power consumption of grinding mill.

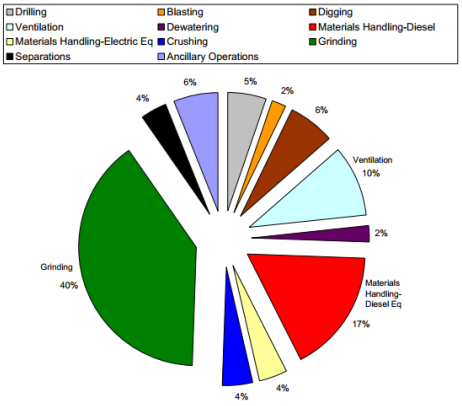


Fig. 1. Contribution of current energy use by equipment across the mining industry. [2]

II. THE DISCRETE ELEMENT METHOD

The discrete element method is a numerical modeling technic that allows to describe the mechanical behavior of distinct materials which interact with their nearest neighbor through local contact laws. DEM uses two simple theories: Newtons Second Law and a force displacement law. The force displacement law calculates forces at the contacts between particles, and then the effect of these forces on the each particle is determined from Newtons Second Law [7]. The update of the position of each particle are then used to calculate the new contact forces and this cycle is repeated for each time step.

Philbert Muhayimana, Department of Mechanical Engineering, JKUAT (corresponding author to provide phone:+2540703283287; e-mail: philbert.muhayimana@students.jkuat.ac.ke).

James K Kimotho, Department of Mechanical Engineering, JKUAT (e-mail: jkuria@eng.jkuat.ac.ke).

Hiram M Ndiritu, Department of Mechanical Engineering, JKUAT (e-mail: hndiritu@eng.jkuat.ac.ke)

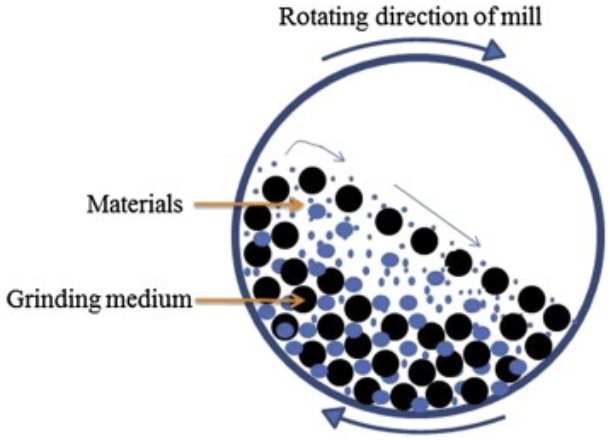


Fig. 2. Illustration of ball mill grinding mechanis [3]

Thus these two contact laws are used to trace the movement of the particles This process is summarized in Fig. 3.

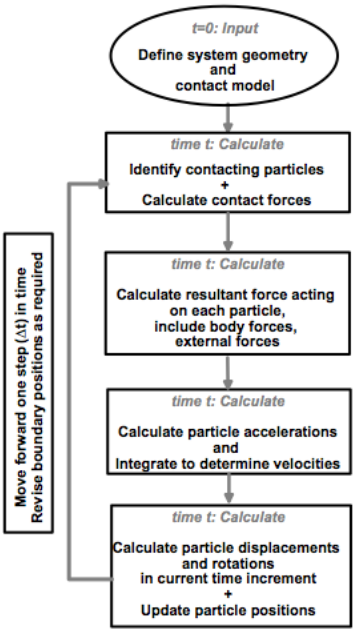


Fig. 3. DEM Calculation Cycle

Cundall and Strack [7] were the first to use the discrete element method, the method was based on the use of an explicit numerical scheme in which the interaction of the particles materials can be monitored contact by contact and the motion of each particles can be modeled. Since then DEM has been adapted to suit many other applications such as, granular flow, powder mixing, and in modeling of many physical systems. Lorig et al [8], analyzed the rock-support interaction, Campbell et al [9], used DEM for granular shear flow analysis, John et al [10], used DEM based on two shaped element to model granular soil behavior, Mishra et al [11], simulated the behavior of balls in a milling machine using DEM, Raasch et al [12], analyzed the trajectories and impact velocities of grinding bodies in planetary ball mill, Datta et al [13] analyzed the power draw in ball mills using the discrete element method, Kim et al [14] analyzed ball movement for research of grinding mechanism of a stirred ball mill with 3D discrete element method.

In DEM particles are usually modeled either in two dimensional (circular discs) or three dimensional (spherical). However, the particles shape can also be modeled in an other shape such as ellipsoid [15], and polygon [16], as well as irregular shape which can be modeled by bonding several spherical or circular particles [17]. Many papers have been published in the literature by using DEM in modelling and simulation of grinding mills, majority of them being limited in 2D. Hlungwani et al. [18] used a 2D laboratory ball mill to validate the DEM modeling of liner profile and mill speed effects. Cleary [19] used DEM to investigate charge behavior and power consumption in relation to operating conditions, liner geometry and charge composition in a 5m ball mill, also limited to the 2D code. Djordjevic et al. [20] have shown that 3D DEM simulations give more accurate results than 2D DEM simulation.

The DEM modeling that is used by many researchers [11], [13] involves determining the particles that are in contact, the amount of overlap and related velocities. Therefore the net forces acting on the contacting force can be obtained. Newton's second law of motion is applied to all particles to determine new particle positions, their velocities and acceleration.

Law of motion is based on the Newton second's law. In a particulate system, a single particle is affected by 3 types of forces: gravity, normal force, tangential forces and its motion can be described as :

$$m_i \frac{dv_i}{dt} = \sum F_{n,i} + F_{t,i} + g, \tag{1}$$

The subscripts i is for representing particle, v is the velocity of the mass center, $\omega$the angular velocity, $F_{n,i}$ the normal force of particle i, $F_{t,i}$ the tangential force of particle i and g is gravity. Taking the time step into consideration, the movement of a particle will be described by five factors: its position x, velocity$x^{'}$, acceleration $x^{"}$, angular velocity ω, angular acceleration $\omega^{'}$. They are all determined by the resultant force and resultant moment.

III. Contact Model

There are two major types of contact models: particle-particle and particle-geometry. The particle-particle contact is the focus in this research, which is mostly used for simulating different materials. Particle-particle contact models, can be classified as contacting force models and non-contacting force models. Generally four types of contact models are: continuous potential models, elastic model, visco-elastic and plastic models, in which the first one belongs to non-contacting force model, while the other three belong to the contacting force model. The continuous potential model is widely applied in molecular systems, which includes van der Waals forces, electrostatic forces, and liquid bridge forces. The elastic model can be sub-classified as linear elastic and nonlinear elastic models.

In DEM, a collision can be either particle to particle or particle to geometry, and is represented by the contact model which show how the colliding bodies interact. Numerous contact models have been used to model the interaction of different particles. The linear spring-and-dashpot used by Cundall and Strack [7], Morrison and Cleary [21], and Datta and Rajamani [13]; the modified linear viscous damping model [22], the bi-linear [22]; Hertz-Mindlin non-linear spring-and-dashpot, [17], [23]. Among all the contact model two model are commonly used.

A. Linear-Spring contact Model

The simplest model uses a linear assumption stating that the displacement is directly proportional to the force and based upon the work of Cundall and Strack. For collisions, the force is decomposed into normal and tangential forces with separate spring-dashpot elements, illustrated in Fig. 4. For each component, we define a spring and dashpot to calculate the force. The transition between a sticking and sliding collision is controlled by the coefficient of friction, which poses an upper limit to the tangential force. Here the characteristic impact (or overlap velocity) is a required input parameter [24].

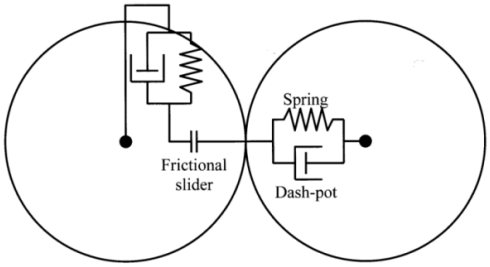


Fig. 4. Linear-spring contact model [25]

In linear spring-and-dashpot contact model, the contact force in normal direction F_n can be governed by

$$F_n = -b_n V_n + k_n U_n, \tag{2}$$

where b_n, U_n, k_n, V_n are The normal damping constant, the overlap of contacting particles, the normal contact stiffness, and the relative normal velocity of particles.

$$F_s = {F_s}^o + k_s(U_s + {U_s}^o), \tag{3}$$

where ${F_s}^o, U_s, k_s, {U_s}^o$ are the contact shear force during the previous time step, the relative tangential displacement, the shear contact stiffness, and the relative tangential displacement for the previous time step, Δt. The linear spring-and-dashpot model is widely used, especially in modeling particles in fluid, but for dry grinding it has one major problem: it is centrally to the law of physics as a model for particle collisions. In this model the viscous damping is assumed to be maximum as the particles are coming into contact and also as the particles are about to separate. This is not what is expected to happen. Damping should be a minimum when the particles first come into contact and also as the particles rebound. Due to the rather unphysical nature of the linear spring-and-dashpot model hypothesized by Sarracino et al. [22] may be adequate for charge motion and power draw predictions, but it wouldnt generate accurate impact energy spectrum predictions

B. Hertz-Mindlin Contact Model

The Hertz-Mindlin model shown in the Fig. 5 below, is the most commonly used within EDEM simulations [17]. The model uses a spring-dashpot response to normal contact between particles and/or geometry and a Coulomb friction coefficient μ for shear interactions and a second spring-dashpot response to tangential or rolling friction interaction. It provides an alternative, to the more common linear spring-and-dashpot modeling. It illustrates more detailed and realistic the interaction between the two particles A and B than the spring-and-dashpot model [26]. Unlike the linear contact model, in the Hertz-Mindlin contact model the normal spring stiffness, k_n, varies according to the amount of overlap, U_n, between the contacting particles, in accordance with Hertzian contact theory developed by Hertz [27]. The total force between the particles can be divided into normal and tangential forces. Spring and damping components are available for both the forces, friction is available only for tangential component and coefficient of restitution is related to the normal force component.This model calculates the normal and tangential forces using material properties such as the coefficient of restitution, Youngs modulus, Poissons ratio, size and mass It is a non-linear elastic model and is thus well suited to the non-cohesive interactions which are to be used within the computational models [28].

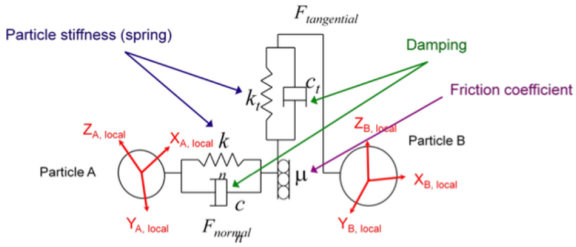


Fig. 5. Hertz-Mindlin model [28]

Using Hertz-Mindlin contact model, the interaction of particles is governed by the Hertz theory

$$F_n = 2/3 P_{max} \pi a^2, \tag{4}$$

where P_{max} is the maximum pressure at the point of contact and a is the area covered by the contacting bodies.

$$a = \frac{3 P_{max} R^*}{4 E^*}, \tag{5}$$

R^* and E^* are the reduced radius of contacting bodies and young modulus respectively. Hence the force- displacement relation in the normal direction is calculated by

$$F_n = -k_n {U_n}^{3/2}, \tag{6}$$

U_n is the contact overlap and k_n is the normal contact stiffness.

IV. Contact Model Parameters

Material properties and material interaction parameters, such as: the spring stiffness - or Youngs modulus and poisons ratio, damping constant or coefficient of restitution and coefficient of friction are required in the early discussed contact models. Chandramohan [29] said that instead of using estimate and approximate values of material interaction properties, they can rather be measured to provide an overall reliable and accurate result in predicting the motion of grinding media in ball mill.

A. Coefficient of Restitution

The damping constant and coefficient of restitution are important interaction parameters for the prediction of charge motion and energy distribution of particles inside the rotating mill. They represent measures of the energy that is lost during a collision. The coefficient of restitution is defined as the ratio of the relative velocities of colliding bodies just before contact, to the relative velocities just after the collision [27].

B. Contact Stiffness, Young's modulus and Poison's ratio

The resultant force from the overlap at the point of contact is a function of the contact stiffness. the selection of this parameter is necessary in DEM measurement. In Hertz-Mindlin model, the stiffness k is known as a function of young's modulus E and Poissons ration ν which are related such that

$$E = 2(1+\nu)G, \tag{7}$$

where G is the elastic shear modulus which is mostly used in DEM simulations.

C. Coefficients of static Friction

The coefficient of static friction is the friction force between two objects when neither of the objects is moving. The coefficient of kinetic friction is the force between two objects when one object is moving, or if two objects are moving against one another.The coefficient of friction governs the initialization of slip between two particles experiencing tangential interaction. Nierop et al. [30] reported the variation of the power draw with coefficient of friction, and suggested that The coefficient of friction may be an important parameter in DEM simulations.

V. Operational parameters that affect grinding mechanism of ball mills

In mineral grinding using ball mills, there are factors that have been investigated and applied in ball milling industries in order to maximize grinding efficiency [5].

A. Mill diameter

Bond [31] observed grinding efficiency as a function of ball mill diameter and established empirical formulas for recommended media and mill speed that take this factor into account. As well, mill with different length to diameter ration for a given power rating will yield different material retention times, the longer the units being utilized for a high reduction ratios and the shorter ones where over-grinding is of concern. Also related to both material and media rentation is discharged.

When designing a ball mill, much consideration is needed on the size of the drum of ball mill and the speed at which the drum is rotating. It has been shown by many researchers that the mill diameter and mill speed have a great impact on the grinding process of granular material [32], [33].

The net power consumption of a milling machine can be calculated using Equation. 8 below. This equation shows that the internal diameter of the milling machine has impact on the net power consumption.

$$NP = \lambda L {D_m}^{2.5} \tag{8}$$

where, NP is the net power, λ is the friction factor, L is the length of the drum, D_m is the internal mill diameter.

Gupta

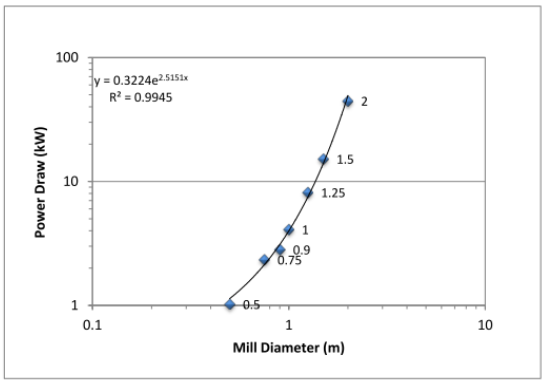


Fig. 6. Predicted Power draw for different diameter mills that have rectangular lifters at a mill speed of 60% CS and 50% mill charge [34]

Rowland also conducted a study and predicted the power draw and charge motion for different mills diameter with the same lifter shapes run at the same operating conditions. 50% ball load was selected according to the previous report that different diameter mills can draw maximum power around this load. Power draw for small scale to large scale mills was then predicted using the DEM simulation. Fig. 6, shows that the power draw for small diameter mills to large diameter mills that as mill diameter increases, simply mill power draw increases [34].

B. Mill speed

When designing a ball mill, much consideration is needed on the size of the drum of ball mill and the speed at which the drum is rotating. It has been shown by many researchers that the mill diameter and mill speed have a great impact on the grinding process of granular material [32], [33]. Deniz [33] investigated the effect of mill speed on the limestone and the clinker samples at batch grinding conditions based on a kinetic model. The effect of operational speed which is the fractional to critical speed φ_c on the grinding for model parameter a_T was found to be different for two different samples: $a_T = 0.0344exp(0.00301\ \varphi_c)$ for

clinker and $a_T = 0.0225exp(0.06183\ \varphi_c)$ for limestone. It was found that, for batch grinding, optimum grinding occurs at $\phi_c = 85\%$. Francioli [5] conducted another study on the effect of operational variables on ball milling. In order to study the effect of mill filling, powder filling, percentage of critical speed, ball size and percentage of solids, he carried out different tests. Grinding media of 25 mm, 30% mill filling, 100% powder filling and 75% of the critical speed was selected as the base condition and all other tests were varied according to the progress of the results and the need to evaluate tests with different operational variable. In this study he has analyzed the effect of mill speed, critical speed also known as the movement of the grinding media adjacent to mill shell during the entire mill rotation was taken as the reference speed. He concluded that ball mills can operate in two distinct regimes depending on the rotation speed: cascade and cataract, as illustrated in Fig. 7.

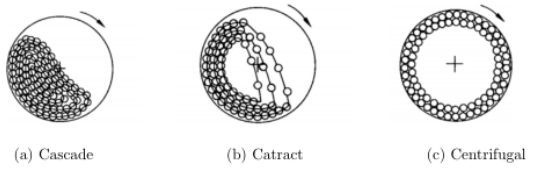


Fig. 7. Distinct regimes of rotation speed of a ball mill [5]

Cascade motion is more likely to result in breakage through attrition whereas cataract would favor collisions and, thus, body breakage. The effect caused by the variation of the critical speed can be seen in Fig. 8, which illustrate that the speed increases the center of mass of the charge inside the mill is dislocated towards the mill wall hence the increase of power consumption. However, when the speed gets closer to the critical speed the center of mass is dislocated to the mill center as the charge starts to centrifuge.

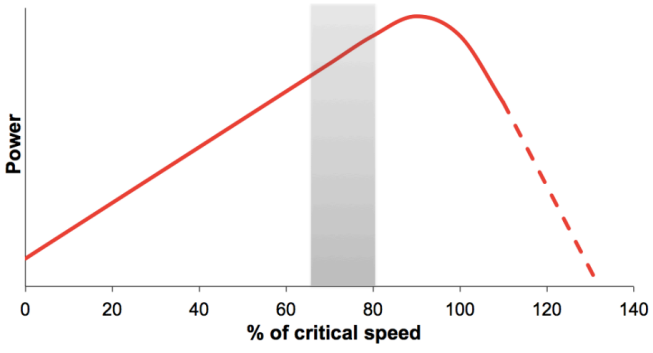


Fig. 8. influence of % of critical speed on power consumption in ball mill [5]

C. Mill filling charge

In grinding, it is also needed to know the rate at which the mill drum volume is occupied by, grinding media and ground material. Mill filling is the percentage of the mill volume occupied by the grinding media and the interstices between them [35]. This operational variable can be written as

$$J = \frac{Vgm}{V_m \times (1 - f_p)}, \tag{9}$$

where, J, is the mill filling level,
Vgm, is the volume of the grinding media inside the mill,
V_m, is the volume of the mill,
f_p, is the fraction volume of interstices between the grinding media usually fp has a value of 0.4.

The charge inside a mill can be given by:

$$f_c = \frac{Vma}{V_m(1 - f_p)}, \tag{10}$$

where Vma is the volume of the material inside the mill [5]. The power consumption of a ball mill can also be calculated using Equation. 11 below

$$P = 2\pi TN, \tag{11}$$

where N is the rotational speed and T is the torque.

The torque necessary to maintain the offset in the center of gravity of the cascading charge from the rest position is given by:

$$T = M_b r_g sin\alpha, \tag{12}$$

where M_b is the ball mass and r_g is the distance between the mill centre and α is the angle of repose of the ball charge.

The calculated power is maximum at about 50% ball load. According to Equation 12. Mill power is a function of ball mass (M_b) and the radius to the center of gravity of ball mass (r_g). As the mill filling increases, the ball mass M_b increases but r_g decreases [13].

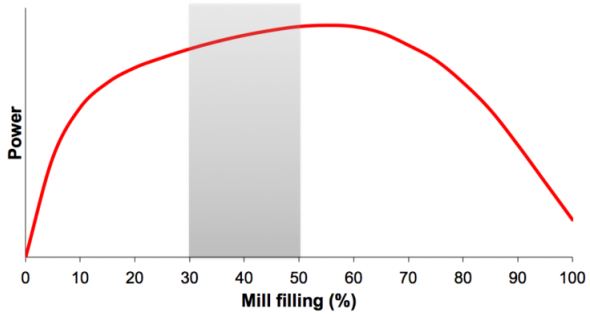


Fig. 9. Effect of mill filling on power consumption of ball mill [5]

Changing the mill filling can also affect the power consumption. Fig. 9, shows that more energy is required when there is an increase of the charge inside the mill as well as, the variation of the center of mass as the percentage of mill filling also plays a major role when it is changed [5].

D. Grinding time

Emami et al. [36] developed a model based on experimental observations to describe the effect of grinding time on the changes of surface area of a rock material during intensive grinding process. Validation and testing the model were performed experimentally using a natural chalcopyrite mineral. The conclusion was that, there is a variation of surface area with respect to time.

It has been reported that in some cases, specific surface area increases first with increasing grinding time but reaches a constant value after a certain grinding time. For such circumstances, Tanaka and Chodacov [36] have proposed the following equation to describe the process of new surface formation:

$$S = S_{max}(1 - e^{k_1 t}) \tag{13}$$

S is the specific surface area at a given time t, and $Smax$ is the maximum attainable specific surface area. The constant k_1 implies the significance of rate constant of the new surface formation.

E. Media size

Grinding media size and shape has a great impact on the grinding operation cost and results in huge consumption of liner and affect the overall performance of ball mill. The ball size in a mill has a significant influence on the mill throughput, power consumption and ground material size Austin et al. [37], Fuerstenau et al. [23]; [38]. The basic condition, which must be met while grinding the material in a mill is that the ball, while breaking the material grain, causes in it stress which is higher than the grain hardness Many researchers have worked on this problem trying to evaluate the effect of media size on the breakage rate material and on the power consumption of ball mill [38], [39].

Kabezya and Motjotji conducted an investigation to determine the effect of the ball diameter sizes on milling operation. A laboratory size ball mill was used with different ball media sizes of 10 mm, 20 mm, and 30 mm respectively. The material used to perform the experiment was Quartz arranged into 3 mono-sizes namely -8mm+5.6mm, -4mm $+2.8mm$, and -2 mm +1.4mm for the experiment. A mill run having a mixture of the 3 ball diameter sizes was also conducted. It was found that, the 30 mm diameter balls were most effective of the three sizes during the grinding of the 3 mono-size feed material samples. The 10 mm diameter balls were the least effective as minimum particle breakage was observed whereas the 20 mm diameter balls were relatively effective to some extent [40].

Magdalinovic also suggests that larger diameter balls have more energy whereas balls having smaller diameters have less energy. These different energies are however relative to the optimum ball diameter, which differs according to the size of the mill as well as the desired size reduction of the feed material [41], [42].

Kabezya and Motjotji also suggested that the mixture of different size of grinding ball can be used for more efficient ball mill. in their findings mixing the 3 different size of grinding ball showed that, the power draw for the ball combination mill run displays a decreasing trend and thus was the most efficiency with regards to the utilization of power towards particle breakage [40].

Kano et al [43], conducted another study by dry grinding gibbsite powder was in air using a tumbling ball mill with mono-size grinding media ranging from 4.8 to 31.7 mm diameter. The grinding device used in his work was a tumbling ball mill made of stainless steel, whose inner diameter, dM, was 121 mm and length, lM, was 142 mm. Steel mono-size balls, having different diameters, dB, of 4.8, 6.4, 7.9, 10.2, 12.7, 15.9, 19.1, 25.4, and 31.7 mm were prepared as grinding media. The ball-filling ratio, J, was kept constant at apparently 40% and the powder sample was charged at 20% for the mill volume. The mill was started running at the rotational speed, N, under dry atmospheric conditions, and the grinding time, t, was varied from 15 to 180 minutes. The main purpose was to investigate ball size effect on grinding rate. The grinding rate increases with an increase in the rotational speed of the mill, subsequently, it falls around the critical speed. The maximum grinding rate shifts toward higher rotational speed range as the ball size becomes large. Kano concluded that, the grinding rate is proportional to the specific impact energy regardless not only of the mill diameter and ball-filling ratio but also of the mono-size ball diameter.

F. Grinding media motion

In mineral grinding the media motion serves to hit the rock material and break it into small and fine materials. Grinding in ball mill has three distinct regime of rotation as seen in Fig. 7. According to Yi Sun et al [44], the motion state of practical charge (material and grinding media) is too complicated to be described precisely. Some researchers considered the grinding media to behave as a single grinding media, others considered the grinding media as the center of the entire mass which was considered as a rigid body, since these considerations ignored some other factors such as the size and shape of particles, it will surely cause variation between theoretical simulations and experiment.

Yi Sun et al [44], used discrete element method to simulate the motion of grinding media inside the drum of ball mill, which demonstrated that the grinding media motion generates the grinding effect in a cascading motion. The force analysis is crucial for the contact model to analyze media motion by considering grinding media as an individual smooth round sphere [44].

In conclusion, to achieve a great impact of the grinding media, it is required to increase the height of fall of the grinding media which will in turn increase the potential energy of the ball. Therefore much consideration of the mill filling level is needed because the mill filling volume rate has much effect on the motion and the contact impact of the grinding media [44].

G. Lifters and liner

Husni Usman [45] studied the effect of lifters configuration on the efficiency of the tumbling mill. The method used was

simulation using Miitraj software. The study of effect of different lifter configuration and operating parameter on the mill efficiency and performance. The rail lifters was concluded to drew higher energy whereas High-Low lifter (Hi-Lo) improves the efficiency of the mill by approximately 22% [45]. For mills with lower mill charge and higher speed, the power usage for different lifters also showed a small difference. The energy requirements of High (Hi) lifter also varied slightly from the other lifters. Which shows that, the Hi lifter improves in the energy efficiency of the mill by approximately 6.7%.The throughput sizes of the different lifters become finer than at the higher load and lower speed. High (Hi) lifter demonstrated higher rate of breakage rates than the rail and Hi-Lo lifters . The Hi lifter at certain operating conditions would improve both the mill efficiency and breakage rate.

Another study conducted by Augustine B. Makokha and Michael H. Moys [46]–[48] has shown that the liner configuration and lifters has much impact on the breakage rate of the ground material. The study was about retrofitting worn liners with cone-lifters. Experiments were performed in a batch wise mode. In all conducted tests, the quartz material was ground for a total period of 4 min. Assessment of the performance of the three liner profiles under investigation and a comparison of the grinding data was made to which a conclusion was made that the liner profile significantly influences the milling rate and fines production in the mill. Fig. 10, also shows another study conducted by Yin et al, showing that the effect of lifter height has less significant effect on power draw of ball mill [49].

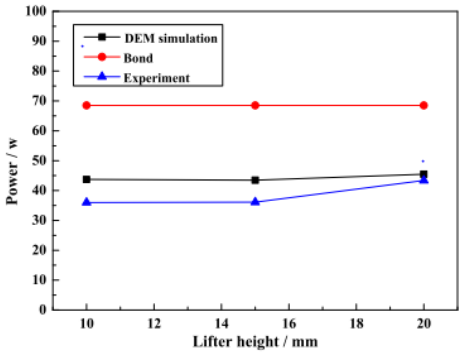


Fig. 10. Power draw at different lifter heights. [49]

VI. Conclusion and Recommendation

From the literature substantial progress has been made in understanding and modeling the grinding process of grinding mills. With the availability of computer power and advanced numerical tools such as DEM grinding mechanism can be predicted using different methods. It was found that Hertz-Mindlin contact model is suitable for modeling the grinding process of a ball mill. It can be able to predict the behaviour of the charge inside the ball mill. It also need a high time step for better prediction. Contact model parameters also needs to be considered as they can affect results in DEM simulations. It was also found that for effective grinding, the media motion needs to be considered. The motion of the charge affect the power consumption of the grinding mill, the quality of throughput material, the wear of the liner and lifter geometry, and the overall efficiency of a grinding mill. Trajectory of grinding media depends on the mill filling level which determines the falling height of the material, the speed of the drum, and the type of liner used. Most of the research considered different parameters and ignored the lifter profile impact on the overall performance of ball mill. The Breakage of the material also was found to be affected by ball diameter and mill speed. The development of the past research done demands further research. Some of the key areas that must be targeted include:

1) First to focus on study of small scale ball mill by considering parameters that affect the grinding process of ball mill, mainly the liner geometry which can also affect the performance of ball mill. In the past researches the main focus has been on the effect of mill filling and grinding media on industrial ball mill with little or no attention on the effect of liner/lifter profile, lifters number, and lifters height, which also has a significant impact on the breakage rate and power consumption of a grinding mill.
2) It is also needed to consider the size distribution or a mixture of ball diameter in a ball mill. Many researchers have shown that bigger grinding balls have more energy whereas balls having smaller diameters have less energy, none has shown the correlation of the grinding ball size with the ground material size. Kano has said that, the grinding rate is proportional to the specific impact energy regardless not only of the mill diameter and ball-filling ratio but also of the mono-size ball diameter. Since there is a contradiction about the effect of ball size or ball diameter on the breakage rate parameter, a deep study is required to establish the correlation of ball size and breakage of the material. Therefore the optimum mixture of balls to mill a given type of charge material will be established.
3) The relationship between: Mill speed, ball size distribution, and lifter geometry, has not yet been clarified. It is not obvious which is dominant in power consumption of a grinding mill. Therefore it is needed to investigate the correlation between mill speed, ball size distribution, and lifters geometry.

References

REFERENCES

[1] D. Saramak, T. Tumidajski, M. Brozek, T. Gawenda, and Z. Naziemiec, "Aspects of comminution flowsheets design in processing of mineral raw materials," *Gospod. Surowcami Miner. / Miner. Resour. Manag.*, vol. 26, no. 4, pp. 59–69, 2010.

[2] D. o. E. US, "Mining Industry energy bandwidth study," pp. 22–30, 2007.

[3] P. Khadka, J. Ro, H. Kim, I. Kim, J. T. Kim, H. Kim, J. M. Cho, G. Yun, and J. Lee, "Pharmaceutical particle technologies: An approach to improve drug solubility, dissolution and bioavailability," *asian journal of pharmaceutical sciences*, vol. 9, no. 6, pp. 304–316, 2014.

[4] K. Asano, H. Enoki, and E. Akiba, "Synthesis of HCP, FCC and BCC structure alloys in the Mg-Ti binary system by means of ball milling," *Journal of Alloys and Compounds*, vol. 480, no. 2, pp. 558–563, 2009.

[5] D. M. Francioli, "Effect of Operational Variables on Ball Milling," *Escola Politecnica*, 2015.

[6] B. Mishra, "A review of computer simulation of tumbling mills by the discrete element method: part icontact mechanics," *International journal of mineral processing*, vol. 71, no. 1-4, pp. 73–93, 2003.

[7] P. A. Cundall and O. D. L. Strack, "A discrete numerical model for granular assemblies," *Géotechnique*, vol. 29, no. 1, pp. 47–65, 1979.

[8] L. Lorig and B. Brady, "13 a hybrid computational scheme for excavation and support design in jointed rock media," in *Design and Performance of Underground Excavations: ISRM SymposiumCambridge, UK, 3–6 September 1984*. Thomas Telford Publishing, 1984, pp. 105–112.

[9] C. S. Campbell and C. E. Brennen, "Computer simulations of granular shear flows," *J. Fluid Mech.*, vol. 151, pp. 167–188, 1985.

[10] J. M. Ti and B. T. Corkum, "Strength behavior of granular materials using discrete numerical modelling," 1988.

[11] B. K. Mishra and R. K. Rajamani, "The discrete element method for the simulation of ball mills," vol. 16, no. August 1991, pp. 598–604, 1992.

[12] J. Raasch, "Trajectories and Impact Velocities of Grinding Bodies in Planetary Ball Mills," *Chem. Eng. Technol.*, vol. 212, no. 1, pp. 245–253, 1992.

[13] A. Datta, B. K. Mishra, and R. K. Rajamani, "Analysis of power draw in ball mills by the discrete element method," *Can. Metall. Q.*, vol. 38, no. 2, pp. 133–140, 1999.

[14] S. Kim and W. S. Choi, "Analysis of ball movement for research of grinding mechanism of a stirred ball mill with 3D discrete element method," *Korean Journal of Chemical Engineering*, vol. 25, no. 3, pp. 585–592, 2008. [Online]. Available: http://dx.doi.org/10.1007/s11814-008-0099-x

[15] J. M. Ting, "A robust algorithm for ellipse-based discrete element modelling of granular materials," *Computers and Geotechnics*, vol. 13, no. 3, pp. 175–186, 1992.

[16] J. Ghaboussi and R. Barbosa, "Three-dimensional discrete element method for granular materials," *International Journal for Numerical and Analytical Methods in Geomechanics*, vol. 14, no. 7, pp. 451–472, 1990.

[17] DEM Solutions, "EDEM 2.4 User Guide," pp. 1–134, 2011.

[18] O. Hlungwani, J. Rikhotso, H. Dong, and M. Moys, "Further validation of dem modeling of milling: effects of liner profile and mill speed," *Minerals Engineering*, vol. 16, no. 10, pp. 993–998, 2003.

[19] P. W. Cleary, "Charge behaviour and power consumption in ball mills: sensitivity to mill operating conditions, liner geometry and charge composition," *International journal of mineral processing*, vol. 63, no. 2, pp. 79–114, 2001.

[20] N. Djordjevic, F. Shi, and R. Morrison, "Determination of lifter design, speed and filling effects in ag mills by 3d dem," *Minerals Engineering*, vol. 17, no. 11-12, pp. 1135–1142, 2004.

[21] R. D. Morrison and P. W. Cleary, "Using DEM to model ore breakage within a pilot scale SAG mill," *Minerals Engineering*, vol. 17, no. 11-12, pp. 1117–1124, 2004.

[22] R. Sarracino, A. McBride, and M. Powell, "Using particle flow code to investigate energy dissipation in a rotary grinding mill," in *Numerical Modeling in Micromechanics via Particle Methods-2004: Proceedings of the 2nd International PFC Symposium, Kyoto, Japan, 28-29 October 2004*. CRC Press, 2004, p. 111.

[23] D. W. Fuerstenau and J. J. Lutch, "The effect of ball size on the energy efficiency of hybrid high-pressure roll mill r ball mill grinding," pp. 199–204, 1999.

[24] C. Kulya, "Using Discrete Element Modelling (DEM) and Breakage Experiments To Model The Comminution Action in a Tumbling Mill," Master thesis, University of Cape Town, 2008.

[25] H. Maghsoodi and E. Luijten, "Chaotic dynamics in a slowly rotating drum," *Revista Cubana de Ftsica*, vol. 33, no. 1, pp. 50–54, 2016.

[26] J. Jaeger, *New solutions in contact mechanics*. Wit Pr/Computational Mechanics, 2005.

[27] K. Johnson, "Contact mechanics, cambridge university press, cambridge.";" 1985.

[28] J. Härtl and J. Y. Ooi, "Experiments and simulations of direct shear tests: porosity, contact friction and bulk friction," *Granular Matter*, vol. 10, no. 4, p. 263, 2008.

[29] R. Chandramohan, "Measurement of particle interaction properties for the incorporation into discrete element methods," Ph.D. dissertation, University of Cape Town, 2005.

[30] M. Van Nierop, G. Glover, A. Hinde, and M. Moys, "A discrete element method investigation of the charge motion and power draw of an experimental two-dimensional mill," *International Journal of Mineral Processing*, vol. 61, no. 2, pp. 77–92, 2001.

[31] F. C. Bond, "Crushing and grinding calculations, part i," *Br. Chem. Eng.*, vol. 6, pp. 378–385, 1961.

[32] R. P. King, "Technical Notes 8 Grinding," *Media*, 2000.

[33] V. Deniz, "The effect of mill speed on kinetic breakage parameters of clinker and limestone," vol. 34, pp. 1365–1371, 2004.

[34] C. A. Rowland, "Bonds method for selection of ball mills," *Advances in Comminution*, pp. 385–397, 2006.

[35] K. K. Kiangi and M. H. Moys, "Particle filling and size effects on the ball load behaviour and power in a dry pilot mill: Experimental study," *Powder Technology*, vol. 187, no. 1, pp. 79–87, 2008.

[36] A. Emami, M. S. Bafghi, J. Vahdati Khaki, and A. Zakeri, "The effect of grinding time on the specific surface area during intensive grinding of mineral powders," *Iranian Journal of Materials science and Engineering*, vol. 6, 2009.

[37] L. Austin, K. Shoji, and P. T. Luckie, "The effect of ball size on mill performance," *Powder Technology*, vol. 14, no. 1, pp. 71–79, 1976.

[38] N. Kotake, K. Daibo, T. Yamamoto, and Y. Kanda, "Experimental investigation on a grinding rate constant of solid materials by a ball mill effect of ball diameter and feed size," vol. 144, pp. 196–203, 2004.

[39] H. Ipek, "Effect of grinding media shapes on breakage parameters," *Particle and Particle Systems Characterization*, vol. 24, no. 3, pp. 229–235, 2007.

[40] K. Km and H. Motjotji, "Material Science & Engineering The Effect of Ball Size Diameter on Milling Performance," vol. 4, no. 1, pp. 4–6, 2015.

[41] N. Nistlaba and S. Lameck, "Effects of Grinding Media Shapes on Ball Mill Performance," Masters thesis, University of the Witwatersrand, Johannesburg, 2005.

[42] N. Magdalinovic, M. Trumic, M. Trumic, and L. Andric, "The optimal ball diameter in a mill," *Physicochem. Probl. Miner. Process.*, vol. 48, no. 2, pp. 329–339, 2012.

[43] J. Kano, H. Mio, F. Saito, and M. Miyazaki, "Correlation of grinding rate of gibbsite with impact energy in tumbling mill with mono-size balls," *Minerals engineering*, vol. 14, no. 10, pp. 1213–1223, 2001.

[44] Y. I. Sun, M. Dong, Y. Mao, and D. Fan, "Analysis on Grinding media Motion in Ball Mill by Discrete Element Method," in *1st Int. Conf. Manuf. Eng. Qual. Prod. Syst.*, vol. I, no. Volume I, pp. 227–231.

[45] H. Usman, "Measuring the efficiency of the tumbling mill as a function of lifter configuration and operating parameters ."

[46] A. B. Makokha and M. H. Moys, "Towards optimising ball-milling capacity : Effect of lifter design," *Minerals Engineering*, vol. 19, pp. 1439–1445, 2006.

[47] A. B. Makokha, M. H. Moys, M. M. Bwalya, and K. Kimera, "A new approach to optimising the life and performance of worn liners in ball mills : Experimental study and DEM simulation," vol. 84, pp. 221–227, 2007.

[48] A. B. Makokha, M. H. Moys, and M. M. Bwalya, "Modeling the RTD of an industrial overflow ball mill as a function of load volume and slurry concentration," *Minerals Engineering*, vol. 24, no. 3-4, pp. 335–340, 2011. [Online]. Available: http://dx.doi.org/10.1016/j.mineng.2010.11.001

[49] Z. Yin, Y. Peng, Z. Zhu, Z. Yu, and T. Li, "Impact load behavior between different charge and lifter in a laboratory-scale mill," *Materials*, vol. 10, no. 8, 2017.

Modal and Harmonic analysis of a small-scale ball mill based on ANSYS

Justin Byiringiro, James K. Kimotho and Hiram M. Ndiritu

Abstract—Ball mill is a grinding machine used in mineral processing, comprising of a rotating cylindrical shell filled with steel balls and the material to be ground. The ball mill is inevitably accompanied by vibration during operation. This vibration may lead to resonance in the structural components if the natural frequency matches the operating frequency. Therefore, it is important to study the dynamic characteristics of the ball mill. Modal analysis is an essential technique which helps in predicting the possibility of resonance by determining the natural frequencies as well as mode shapes. In this study, modal analysis of a small-scale ball mill developed in Jkuat is carried out in ANSYS Workbench 16.0. The first ten modes of vibration are extracted for two scenarios, one, for the ball mill and frame support (drum, shaft and frame support) and two, for the drum and shaft. Harmonic analysis is done to visualize the response of the ball mill structure under dynamic loading. The results show that some frequencies fall within the expected operating frequencies.

Keywords—Ball mill, Finite Element Method, Modal Analysis, Resonance, Harmonic response.

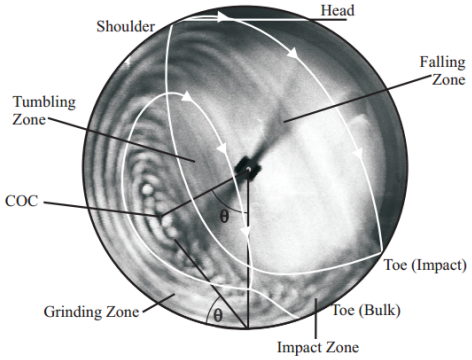


Fig. 1. A typical ball mill motion [2].

I. INTRODUCTION

BALL mill is a machine used in grinding minerals. It is a rotating cylindrical shell filled with steel balls and material to be reduced. Due to the rotation of the drum , the charge is lifted toward the shoulder, the point at which the charge material separates from the mill shell. From the shoulder, the charge falls towards the toe, the zone of intersection where the tumbling charge impacts the material below as demonstrated in Fig. 1. As the charge transits from the toe back to the shoulder, it is subjected to grinding forces. Throughout the cycle, the charge is subjected to both impact and grinding forces. The tumbling of steel balls and the material to be ground are the main source of vibration in a ball mill [1]. This vibration may cause operational problems in ball mill structure even leading to resonance if not well modeled.

Resonance is the tendency of a mechanical structure to absorb more energy when the frequency of its oscillations matches the system's natural frequency of vibration than it does at other frequencies, resonance may cause severe vibration even leading to failure.

Modal analysis is a technique to study the dynamic characteristics of a structure under vibrational excitation [3]. Natural frequencies, modes shapes and mode vectors of a machine can be determined using modal analysis. Modal analysis allows the design to avoid resonance and gives an idea of how the design will respond to different types of dynamic loads. Ball mill is an example of a machine whose dynamic characteristics can be better studied by modal analysis.

Modal analysis since 1960s, has been widely used as an important method in fault diagnosis [4]. A lot of researches on the modal analysis of the ball mill have been conducted. Chen et al. [4] have used ANSYS workbench, to carry out modal analysis of an oversize ball mill tube where first ten natural frequencies were calculated, it was found that resonance could not happen since the lowest natural frequency of ball mill tube was much higher than the operating frequency, also from vibration modes it was found that crack could happen in the tube much easier than in other area since all vibrations modes was in radial direction and the maximum amplitude occurred in the middle of the tube. However this study considered only the ball mill tube. In [5] Porto B.Thiago, Mendonca Q.Beatriz and Carvalho S.G.Lucas used Finite Element Method to establish the basic design requirements for structures subjected to dynamic action, Ball mill and dryer were discussed as case studies. The analysis of the natural frequency of the base of equipments was performed and it was found that there were no risk of resonance since these frequencies was outside the limits of the operating frequency. However the finite element method in this study considered only the base of equipments. Quan [7] explored the tumbling mill resonance using six different kinds of mill FE models. Modal analysis was performed to investigate modes shapes and natural frequencies at different mill rotation speeds in order to investigate resonance behavior. However in this study the simulation model used, only includes the mill tube while the other parts of the mill system was not

Justin Byiringiro, Department of Mechanical Engineering, JKUAT (corresponding author to provide phone:+254(0)703283295; e-mail: justin.byiringiro@students.jkuat.ac.ke).

James Kimotho, Department of Mechanical Engineering, JKUAT (e-mail: jukuria@eng.jkuat.ac.ke).

Hiram Ndiritu, Department of Mechanical Engineering, JKUAT (e-mail: hndiritu@eng.jkuat.ac.ke).

taken into consideration.
Radziszewski [8] investigated the resonance of ball mill by studying the natural frequency. The author study the ball mill as a rotating machinery. Unlike stationary machines, in rotating machineries resonance is related to the rotation speed, because their natural frequencies vary along the operating speed. Fig. 2, shows that the resonant behavior related to mill rotation speed.

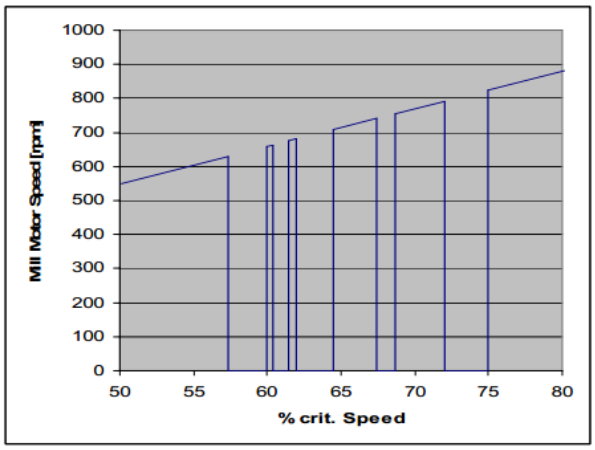


Fig. 2. Mill resonance behavior related with mill rotation speed [8].

The objective of this study is to determine natural frequencies, study the mode shapes of the ball mill and subject to a harmonic loading in order to predict the dynamic characteristics of a ball mill and to avoid possible resonance. Modal analysis and harmonic response are carried out in Finite Element Analysis ANSYS workbench 16.0. Natural frequencies are extracted for two conditions; first for ball mill drum with the frame support and for drum with shaft. The geometric modeling of ball mill was done in SOLID WORKS.

II. Finite Element Modeling

A. Theory of Modal and Harmonic response analysis

Finite element calculation mode is structural dynamics of the eigenvalue. Eigenvalues and eigenvectors is natural frequencies and mode shapes modal analysis. The dynamic equation of motion [9], is shown in Equation 1.

$$[m]\{\ddot{u}\} + [c]\{\dot{u}\} + [K]\{u\} = \{F(t)\} \tag{1}$$

where; $[m]$ is mass Matrix, $[c]$ is damping matrix, $[K]$ is stiffness matrix, $\{\dot{u}\}$ is nodal velocity, $\{\ddot{u}\}$ is nodal acceleration, $\{u\}$ is nodal displacement, $\{F(t)\}$ is excitation Force.

The eigenvalue problem is solved with undamped modal analysis as shown in Equation 2.

$$[m]\{\ddot{u}\} + [K]\{u\} = \{0\}. \tag{2}$$

The free vibration mode of the structure is harmonic vibration, so the displacement is a sine function.

$$X = A\sin(\omega t). \tag{3}$$

Combining Equation 3 and 2 result in:

$$([K] - \omega^2[M]) = 0. \tag{4}$$

Equation 4 represents an eigenvalue problem , where ω_i^2 is eigenvalues and ω_i is natural circular frequency.

Modal analysis is actually solving eigenvalue and eigenvector.

III. Modal and Harmonic analysis

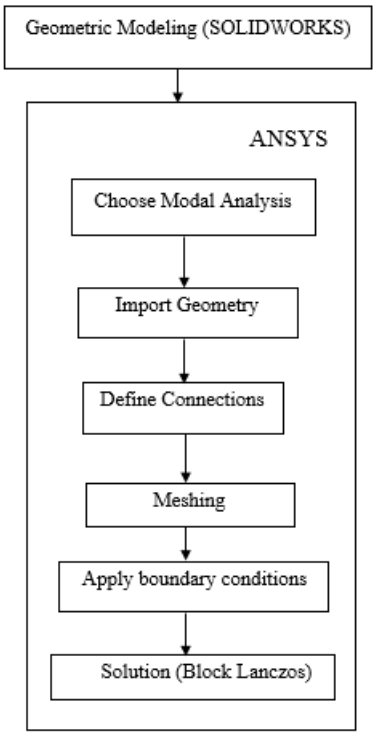


Fig. 3. Flow chart of procedure used in modal analysis

A simplified 3D model of ball mill developed in jkuat is modeled in SOLIDWORKS then imported to ANSYS workbench 16.0 using Initial Graphic Exchange Specification (IGES) file extension. Procedure used in modal analysis is illustrated in Fig. 3.
Modal analysis is done for two conditions;

1) Ball mill and frame support
2) Drum and shaft, considering the rotation velocity.

For condition 1 modal analysis of ball mill and frame support as shown in Fig. 4, is done for predicting the dynamic characteristic of the ball mill tube together with the frame support. For condition 2 the rotor dynamic analysis (modal analysis) of only the drum and the shaft as shown in Fig. 6, is performed in order to check whether the drum and the shaft are safe from resonance caused by the whirling of the shaft.
In operation, the ball mill is subjected to two major excitation frequencies, that is, the excitation frequency from the comminution process and the rotation frequency of the drum. For now, only the rotation frequency has been considered in

analysis as the modeling of the internal process of the ball mill to get the excitation frequency from the comminution process is still continuing.

A. *Condition 1: Modal analysis of ball mill with frame support*

Modal analysis of ball mill structure is done to obtain the first ten natural frequency and vibration mode. There are 182725 nodes and 140385 elements generated in total in the model of the structure. A simplified model used in modal analysis is shown in Fig. 4.

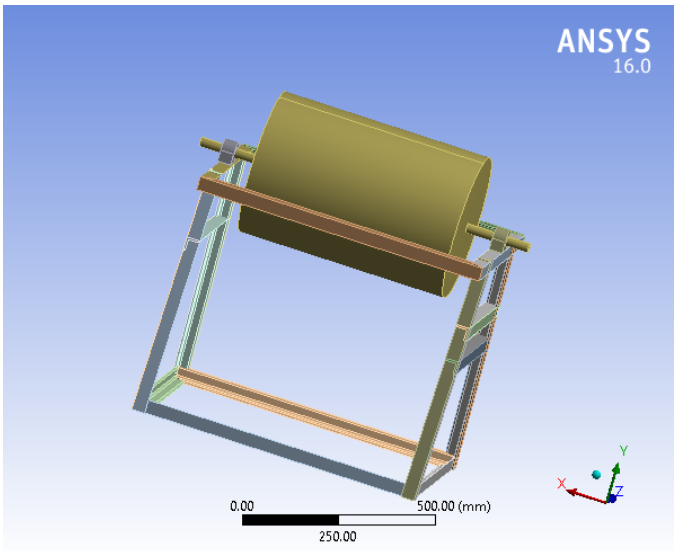


Fig. 4. A simplified model of Ball mill from jkuat

1) Mesh Design: The model is divided into 182725 nodes and 140385 Elements. The mesh used and quality evaluation parameters are shown in Table I.

TABLE I
MESH QUALITY EVALUATION PARAMETERS

No	Parameters	Averages Value Used	Evaluation Standard and Value Range
1	Aspect Ratio	1.6522	Optimum value is 1 Warning value is 20
2	Element Quality	0.85032	ranges from 0 to 1 Optimum value is 1
3	Jacobian Ratio	1.0532	Optimum value is 1 Warning value is 40
4	Wraping Factor	2.57E-03	Optimum value is 0 limit is 7
5	Parallel Deviation	5.7603	Optimum value is 0 Warning value is 70
6	Maximum Corner Angle	96.835	Optimum value is 90 Warning value is 155
7	skewness	0.19596	ranges from 0 to 1 Optimum value is 0
8	Orthogonal Quality	0.89292	ranges from 0 to 1 Optimum value is 1

The element quality and skewness are the most important parameters to evaluate the mesh quality. From Table I the element quality value is 0.85032 and skewness is 0.19596. The mesh quality is excellent when the skewness value is less than 0.25 [10]. Therefore, the mesh quality are good for the model calculation.

2) Boundary conditions: While carrying out modal analysis of the ball mill, it is only needed to set constraints without considering the force. For the boundary condition in the modal analysis, the base of the frame support is fixed. Fig. 5, shows the boundary conditions applied.

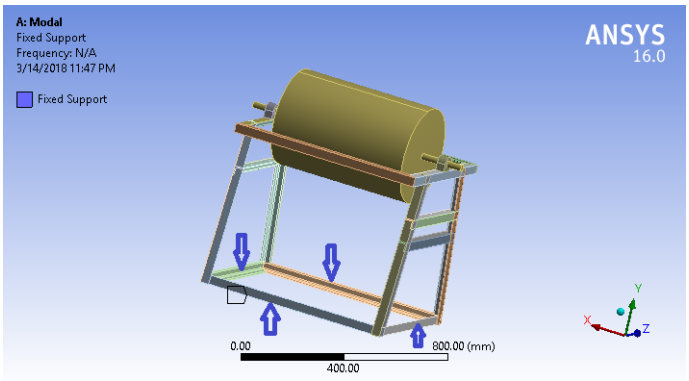


Fig. 5. Boundary conditions applied

3) Solution scheme: According to the vibration theory, the lower order modes have a huge impact on vibration [10]. Therefore the first 10 natural frequencies and vibration modes of the ball mill are solved by Block Lanczos method. The advantage of this method is that the mesh quality is low and the running speed is fast [11]. Settings used in ANSYS are shown in Table II and Table III.

TABLE II
ANSYS SETTINGS USED

object name	**Modal (A5)**
State	Solved
Physics Type	Structure
Analysis Type	Modal
Solver Target	Mechanical APDL
Environment Temperature	22 oC
Generate Input Only	No

TABLE III

Object Name	Pre-stress (None)
State	Fully Defined
Pre-Stress Environment	None

B. *Condition2: Ball mill drum and shaft*

For this analysis, only the drum and shaft are analyzed as shown in Fig. 6. The drum is simplified to simply supported constrained beam and restricted in the ad-axial surfaces which lie in the shaft at the ends of the roller. It was assumed that the total mass of the drum is a point mass acting at the center of the drum which is 200kg (mass of the drum with the grinding medium).

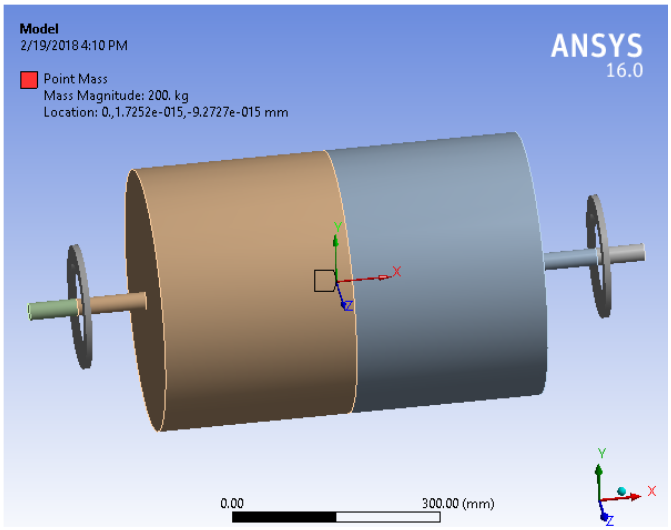


Fig. 6. Model used for rotor Dynamic analysis

Harmonic response analysis is used to predict the sustained dynamic behavior of the structure under predetermined excitation, verifying whether or not structure will successfully overcome harmful effects of forced vibration. In harmonic response analysis the constraint is as same as modal analysis since they share some informations as shown in Fig. 7. The frequency range is set to 0-200Hz. In a harmonic analysis, the peak response will correspond with the natural frequencies of the structure. The analysis settings used for harmonic response are shown in Table IV.

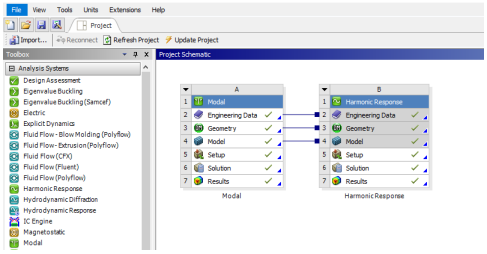


Fig. 7. Setting harmonic response

TABLE IV
ANALYSIS SETTING FOR HARMONIC RESPONSE

State	Fully Defined
Range Nininum	0.Hz
Range Maximum	200.Hz
Solution Intervals	80
Solution Method	Mode Superposition
Cluster Results	No
Modal Frequency Range	Program Controlled

IV. RESULTS AND DISCUSSIONS

Modal analysis determines the vibration characteristics of the structure, by determining the mode shapes and natural frequencies. First ten natural frequency results are obtained between 0-150 Hz range. Six of the ten mode shapes calculated are illustrated in Fig. 9, Fig. 10, Fig. 12, Fig. 13 and Fig. 14. Mode shapes frequencies and characteristic are tabulated in Table V.

A. Modal analysis results for condition 1

TABLE V
FREQUENCIES AND CORRESPONDING VIBRATION MODES

Mode	Frequency (Hz)	Type of mode
1	3.6	Bending along X axis
2	3.6562	Bending along X axis
3	9.152	Bending along Z axis
4	19.655	Bending along Y axis
5	19.788	Twisting along X axis
6	39.576	Bending along Y axis
7	70.911	Bending along X axis
8	121.95	Bending along X axis
9	128.5	Bending along Z axis
10	146.45	Bending along Y axis

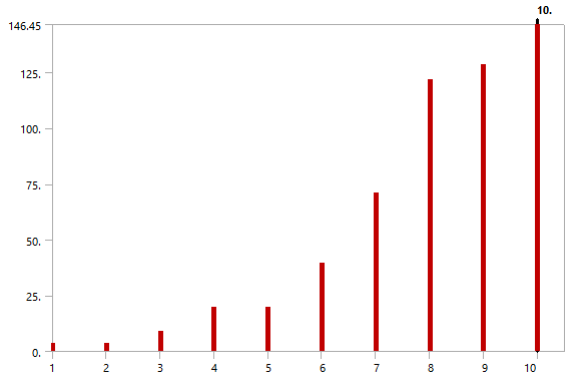


Fig. 8. Variation of number of modes vs frequency. X-axis contains number of modes and Y-axis contains frequency

B. Mode shapes

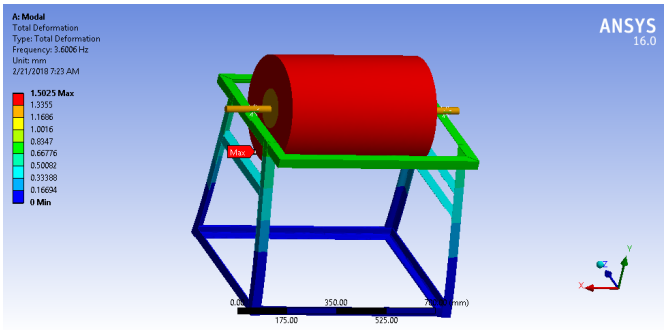


Fig. 9. The 1st mode

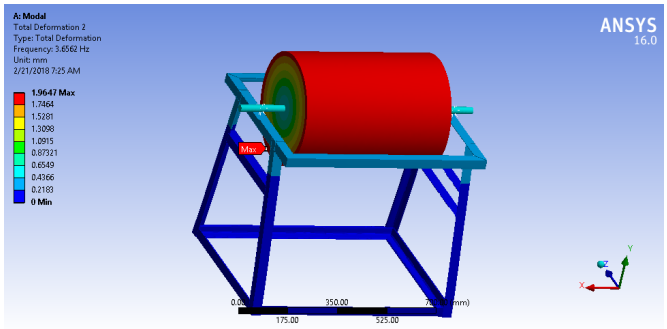


Fig. 10. The 3rd mode

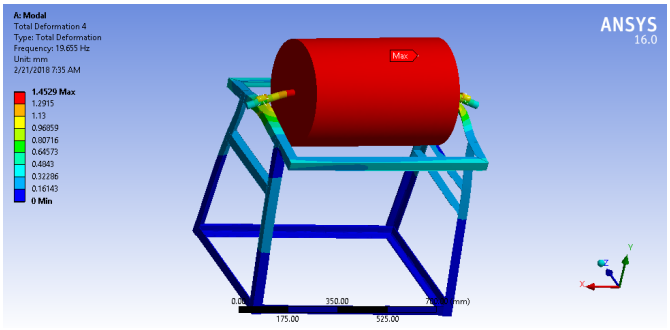


Fig. 11. The 4th mode

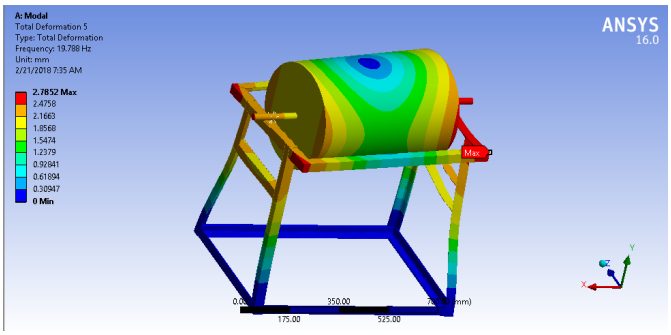


Fig. 12. The 5th mode

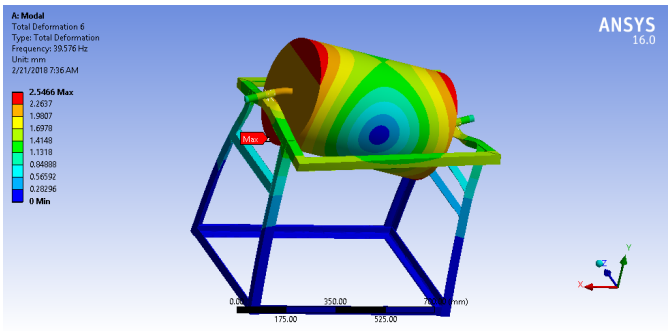


Fig. 13. The 6th mode

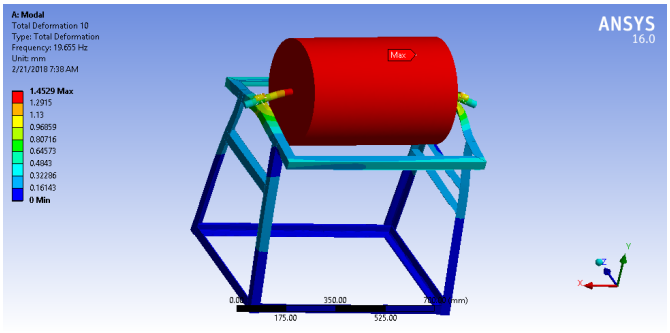


Fig. 14. The 10th

As shown in Table V, the first and second order natural frequencies are close and their vibration mode is bending along X axis, with maximum deformation located on the drum. The vibration mode of the third order natural frequency is bending along Z axis with maximum deformation on part of frame support and part of drum. The fourth and fifth order natural frequencies are also close but their modes of vibration are different, the vibration of the fourth is bending along Y axis with the maximum deformation on the drum, vibration mode of the fifth is twisting along X axis with the maximum vibration on the upper part of the frame support. The mode of vibration of the sixth order natural frequency is bending along Y axis with maximum deformation on the drum. The seventh and eighth order natural frequencies are the same as the first and second. The ninth vibration mode is the same as the third. The tenth vibration mode is the same as the fourth.

From modal analysis shows that the first natural frequency is higher than the operating frequency of the ball mill, as said above only the rotation frequency is considered. The first natural frequency which is responsible for the resonance vibrations is equal to 3.6006Hz and is not close to the rotation frequency of the drum. The operating speed of ball mill is 50rpm, that is 0.83Hz. However for the fourth, sixth and tenth order vibration modes as shown in Figs (11 , 13, 14), it obvious that the whirling of shaft due to shaft deflection caused by loading of the drum, introduces some vibration which could have a high amplitude.

C. Modal analysis results for condition 2

Damped frequency (Hz), stability (HZ), modal damping ratio and logarithmic decrement are illustrated in Fig. 15.

Set	Solve Point	Mode	Damped Frequency [Hz]	Stability [Hz]	Modal Damping Ratio	Logarithmic Decrement
1.	1.	1.	9.834e-004	3.2577e-011	-3.3127e-008	2.0814e-007
2.		2.	6.2216e-003	-2.773e-009	4.4571e-007	-2.8005e-006
3.		3.	20.247	-0.37175	1.8358e-002	-0.11537
4.		4.	20.34	-0.37701	1.8532e-002	-0.11646
5.		5.	86.108	-6.9027	7.9907e-002	-0.50368
6.		6.	86.451	-6.6066	7.6197e-002	-0.48016
7.		7.	0.	-200.14	1.	N/A
8.		8.		-204.07		
9.		9.	302.63	-29.084	9.5663e-002	-0.60384
10.		10.	306.49	-29.822	9.6844e-002	-0.61136
11.	2.	1.	9.834e-004	3.1092e-010	-3.1617e-007	1.9865e-006
12.		2.	6.2216e-003	-2.7317e-009	4.3906e-007	-2.7587e-006
13.		3.	20.246	-0.37169	1.8356e-002	-0.11535
14.		4.	20.341	-0.37707	1.8534e-002	-0.11648
15.		5.	74.044	-5.7744	7.775e-002	-0.49
16.		6.	100.54	-7.704	7.6405e-002	-0.48147
17.		7.	0.	-200.21	1.	N/A
18.		8.		-204.06		
19.		9.	302.63	-29.084	9.5663e-002	-0.60384
20.		10.	306.49	-29.822	9.6844e-002	-0.61137

Fig. 15. Damped frequency (Hz), stability (Hz), modal damping ratio and logarithmic decrement

Fig. 16, shows the frequency at each calculated mode

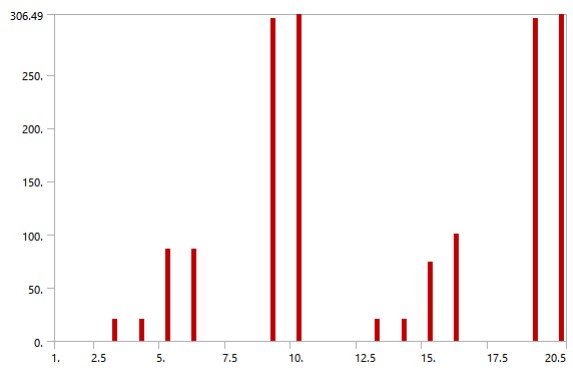


Fig. 16. Frequency at each calculated mode

D. Modal shapes

The first ten order modes extracted using Block Lanczos method, of the ten modes calculated five modes are illustrated in Fig. 17, Fig. 18, Fig. 19, Fig. 20 and Fig. 21.

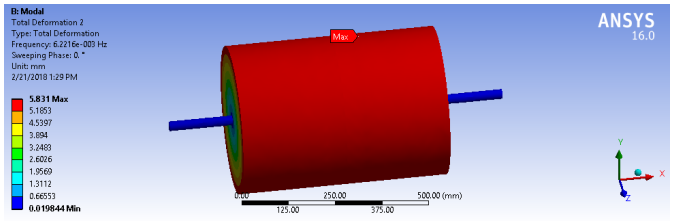


Fig. 17. The 2nd mode

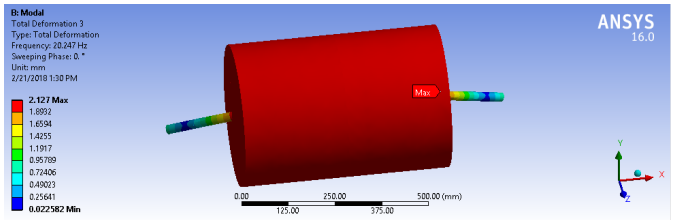


Fig. 18. The 3rd mode

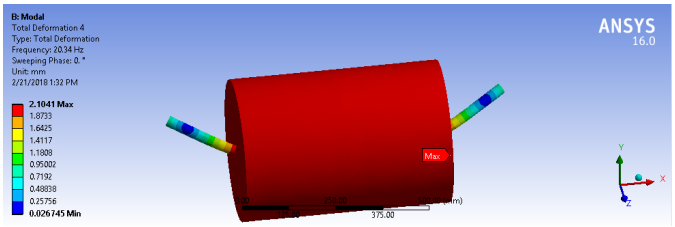


Fig. 19. The 4th mode

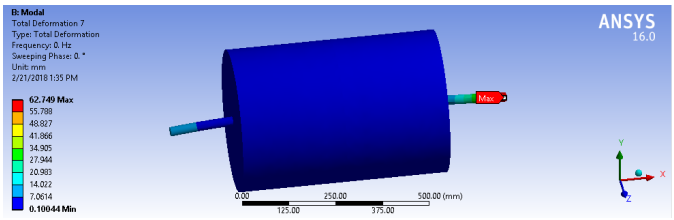


Fig. 20. The 7th mode

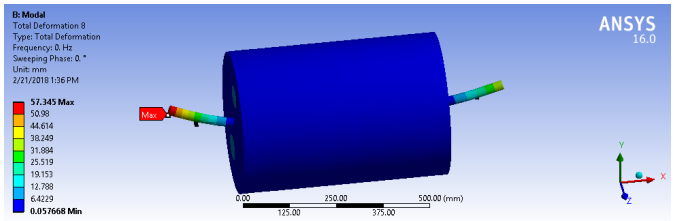


Fig. 21. The 8th mode

E. Campbell diagram

The Campbell diagram is one of the most important tools for understanding the dynamic behavior of the rotating machines. The Campbell diagram used in this research consists of a plot of the natural frequencies as a functions of the rotation velocity.

Fig. 22, shows the Campbell diagram of natural frequencies (Hz) as a function of rotation velocity (rad/s).

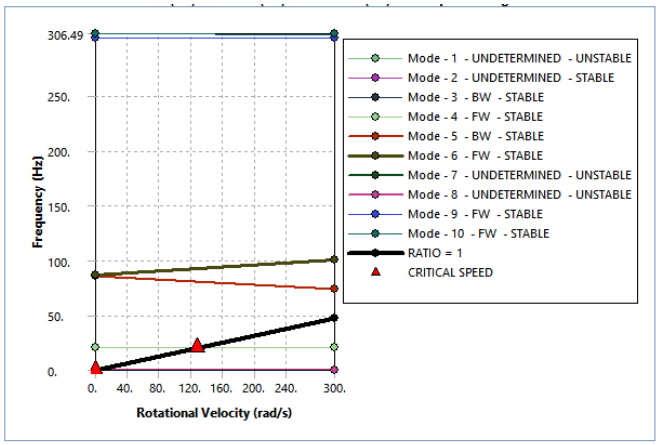


Fig. 22. Campbell diagram (a)

Mode	Whirl Direction	Mode Stability	Critical Speed	0. rad/s	300. rad/s
1.	UNDETERMINED	UNSTABLE	NONE	9.834e-004 Hz	9.834e-004 Hz
2.	UNDETERMINED	STABLE	3.9092e-002 rad/s	6.2216e-003 Hz	6.2216e-003 Hz
3.	BW	STABLE	127.21 rad/s	20.247 Hz	20.246 Hz
4.	FW	STABLE	127.8 rad/s	20.34 Hz	20.341 Hz
5.	BW	STABLE	NONE	86.108 Hz	74.044 Hz
6.	FW	STABLE	NONE	86.451 Hz	100.54 Hz
7.	UNDETERMINED	UNSTABLE	NONE	0. Hz	0. Hz
8.	UNDETERMINED	UNSTABLE	NONE	0. Hz	0. Hz
9.	FW	STABLE	NONE	302.63 Hz	302.63 Hz
10.	FW	STABLE	NONE	306.49 Hz	306.49 Hz

Fig. 23. Campbell diagram (b)

The highest deflections are 62.749mm and 57.345mm on seventh and eight modes respectively as shown in Fig. 20, and Fig. 21.

Critical speeds are 0.039092 rad/s , 127.21 rad/s and 127.8 rad/s as shown in Fig. 23. The first critical speed is within the operating range of the ball mill.

F. Harmonic response results

Harmonic analysis result is used to verify the steady-state response of a structure, which enable researches and ball mill designers to determine whether the ball mill can withstand resonance or other structure problems related to vibration during its operating life. Harmonic analysis calculates the response of a structure to cyclic loads over a frequency range and plot the response of a structure on amplitude versus frequency graph. Post-processing option in ANSYS is used to plot the amplitude versus frequency graph as shown in Fig. 24.

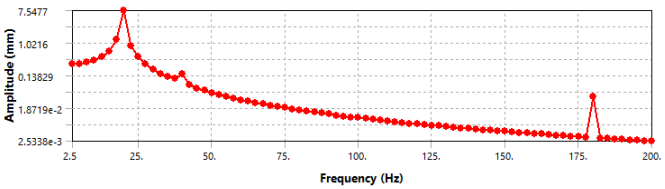


Fig. 24. Variation of displacement amplitude with different exciting frequencies

Fig. 24, shows that the maximum amplitude is generated in 20Hz and the corresponding value of amplitude is 7.5477mm. Resonance occurred at this frequency since it generate the highest amplitude. This is close to the 4th and 5th order predicted natural frequencies. Therefore the 4th and 5th modes shapes are the resonance frequency for ball mill. The drum and the shaft will be the most affected since the highest amplitude of the 4th and 5th natural frequencies are located there. The structure of the ball mill needs to be optimized in order to enhance the stiffness.

V. Conclusion

This study presents modal and harmonic analysis of small-scale ball mill developed in Jkuat. The first ten natural frequency and corresponding vibration mode of two conditions are determined. Campbell diagram has been plotted graphically and critical speeds were obtained. The vibration of number of modes versus frequency has been plotted graphically. The description of mode with the corresponding frequency has been tabulated. Harmonic response of the ball mill for the excitation in the range of 0-200Hz has been studied. Variation of amplitude with respect to frequency has been graphically plotted. From modal the above analysis, the following Conclusions are drawn:

1) The structure of the ball mill need to be optimized especially the ball mill drum and the shaft, in order to avoid not only resonance but also vibrations which can cause higher deflection.
2) In working condition it is advised to avoid the above natural frequencies and critical speeds of the ball mill in order to prevent resonance.
3) The modal analysis can provide reference for the design of ball mill and the selection of reasonable parameters.

References

[1] T. Q. Li, Y. X. Peng, and Z. Zhu, "Energy dissipation during the impact of steel ball with liner in a tumbling ball mill," *Material research innovations*, vol. 19, 2015.
[2] S. Martins, "Exploring Tumbling Mill Dynamics Through Sensor Development," Ph.D. dissertation, McGill, 2011.
[3] J. He and Z.-F. Fu, *Modal Analysis*, 1st ed. Oxford: Butterworth Heinemann, 2001.
[4] L. Chen, Y. Xue, Y. Liu, and J. Li, "Modal analysis of oversize ball mill tube based on ANSYS," vol. 890, pp. 62–65, 2014.
[5] P. B. Thiago, C. S. G. Lucas, and M. Q. Beatriz, "Basic Design Requirements for Structures Subjected to Dynamic action," vol. XXXI, pp. 13–16, 2012.
[6] H. R.M., D. S.J., and B. Stone, "An Experimental investigation into Torsional vibration in ball mills," in *Fifth International Congress on sound and vibration*, 1997.
[7] Y. Quan, "Finite element analysis of tumbling mill design and operating effects on liner bolt stresses, liner stresses and mill resonance," Maters, McGill, 2006. [Online]. Available: http://journals.cambridge.org/abstract_S0165115300023299
[8] P. Radziszewski, Y. Y. Quan, and J. Poirier, "Design Parameters affecting Tumbling Mill Natural Frequencies," no. 12 m.
[9] S. S. Rao, *Mechanical Vibrations*, 5th ed., Q. Tacy, Ed. Pearson, 2010, vol. 67.
[10] F. Hao, Z. Yong-jun, and W. Xiao-ping, "Dynamic Characteristics Analysis and Structure Optimization Study of Glaze Spraying Manipulator," vol. 02005, 2016.
[11] D. Yang, Z. Yu, L. Zhang, and W. Cheng, "Modal Analysis of Automobile Brake Drum Based on ANSYS Workbench," vol. 75, pp. 608–612, 2017.

Towards Device Driven 5G: Radio Resource Allocation Perspective

Njiraine Morris M., Kibet Langat and Stephen Musyoki

***Abstract*—**5G cellular network is the next revolutionary wireless network system that is anticipated in the next few years with exceptional capabilities to achieve the IMT 2020 requirements. The 5G network will introduce a high level of flexibility, user centric, ultra reliable and low latency communications with a plethora of novel applications. Introduced in the LTE advanced system, carrier aggregation technology was one of the major milestone in radio resource management, significantly increasing the overall network performance.

5G will be a collision of networks tiers of various transmit powers, sizes, different radio access technologies, backhaul connections, and unprecedented number of heterogeneous and smart devices. Device to Device (D2D), Massive Machine Communication (MMC), enhanced carrier aggregation along with other technologies proposed in 5G such as energy harvesting and wireless network virtualization will introduce complexity of the resource scheduling.

While the network design is envisioned to overcome the fundamental challenges of the existing and previous mobile networks, there are some clearly unprecedented challenges that can be foreseen. A survey of recently proposed resource allocation strategies for 5G, design aspects, issues and the likely challenges to be encountered in the resource allocation for the multi-tier architecture of this device driven network are presented.

***Keywords*—**5G, IMT 2020, Radio Resource, Allocation Strategies, Future Networks, D2D Communication

I. Introduction

In the recent days, the world has demonstrated considerable dynamism in the usage of wireless and cellular communication through development trends in emerging applications and rapid demand of high speed internet, with massive interconnected wireless devices such as machines, smart cities, sensors and tablets. Fifth generation (5G) mobile network is expected to revolutionise the user experience, giving infinite capacity to the user, as opposed to the previous networks that focussed primarily on transmission efficiency of particular services such as voice and video streaming. The mobile networks will be expected to sustain manifold wireless requirements in different use cases under 5G. To sustain these emerging applications with diverse traffic characteristics, an entire engineering paradigm shift is vital in the development of the 5G mobile network.

Through Internal Mobile telecommunications 2020 (IMT-2020) also known as 5G, the International Telecommunications Union Radiocommunications Standardisation Sector (ITU-R) outlined manifold design goals for user experienced data rate, peak data rate, connection density, area traffic capacity, mobility and latency, with potential for future evolution [1]. In achieving these goals, a sufficient 5G radio access network system will be required, having to integrate the revolved and evolved multiple radio access technologies of previous mobile network generations. In meeting the requirements and the needs of 5G service landscape, three main 5G Service types, also known as scenarios are considered.

Massive Machine to machine (M2M) communications (mMTC) will be offered in the 5G New Radio (NR), requiring wireless connectivity for up to tens of billions of devices in the world that are network enabled. Here, wide area coverage, scalable connectivity for the growing number of devices per cell and deep indoor are the main priorities.

Extreme Mobile BroadBand (xMBB), also known as enhanced mobile broadband (eMBB), demands both low latency communication and extremely high data rates, and reliable broadband access over large coverage areas.

Ultra-reliable Machine type communications (uMTC), also known as ultra-reliable and low latency communications (URLLC), related to mission critical industry control, vehicle to anything (V2X) communications, and smart grid, among others.

The Fig. 1 indicates the three key service types or scenarios introduced for 5G network, together with the requirements for the scenarios in terms of performance parameters such as latency, area traffic capacity and connection density.

Depending on services offered and propagation conditions best suitable access technology (or combination of access technologies in the case of coordinated Multi points CoMP), multiple connectivity or load sharing can be assigned through optimised selection schemes of multiple radio access technology [2]. In integrating all these technologies and features in 5G, agile radio resource management is required to ensure optimised network performance is achieved. To achieve low latency and ultra-reliable communication in 5G system, we need to understand the tools appropriate for design of radio resource allocation under the limitations of availability, reliability and delay. On the other hand, D2D communication introduced in 5G offers various resource management problems [2] [3] [4] [5]. This research paper first introduces 5G architecture briefly, enabling technologies, and then carries out a survey of the studies done recently on resource allocation in D2D enabled 5G networks.

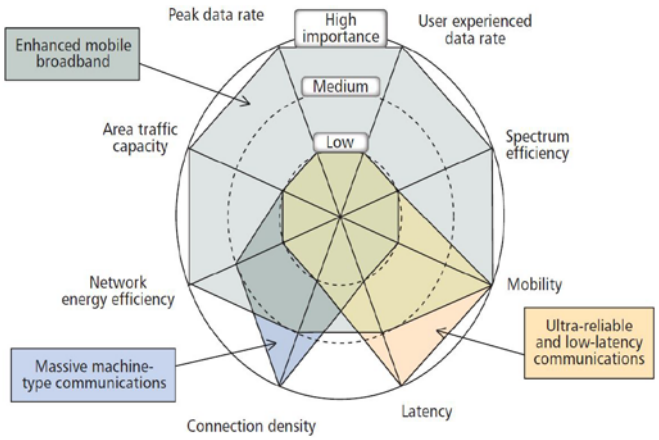


Fig. 1: 5G Case Scenarios with appropriate performance metrics

II. 5G RAN Architecture: Overview

The new RAN consists of gNBs offering the NR C-plane and the U-plane protocol terminations towards the user equipment (UE) [7]. The RAN also consists of LTE eNBs that offer E-UTRA C-plane and U-plane protocol terminations towards the UE, and whether to establish a new RAN logical node that can provide both the E-UTRA and NR C-plane and U-plane protocol terminations towards the UE will be decides in the normative phase. Xn interface interconnects the logical nodes in the new RAN, which are then connected to the NGC through the NG interface as shown in Fig 2. Many-to-many relation is supported by the NG interface between the logical nodes in the New RAN and the NG-CP/UPGWs [7].

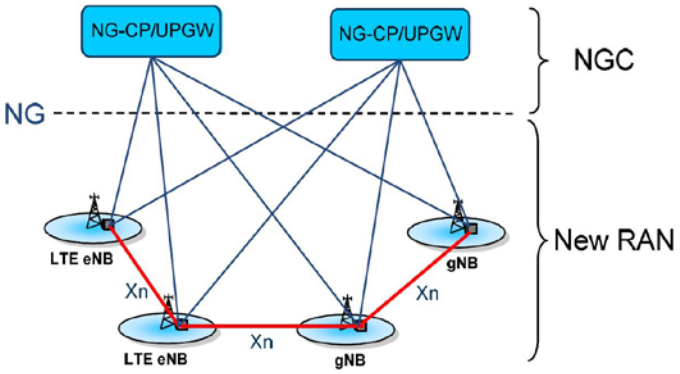


Fig. 2 5G RAN Architecture [6]

The RAN and NextGen Core (NGC) ensures quality of services by mapping packets to appropriate Data Radio Bearers (DRBs) and QoS flows, as shown in Fig. 3. Therefore, there is a 2-step mapping of IP flows to QoS flows (NAS) and then to DRBs from the QoS flows (access stratum).

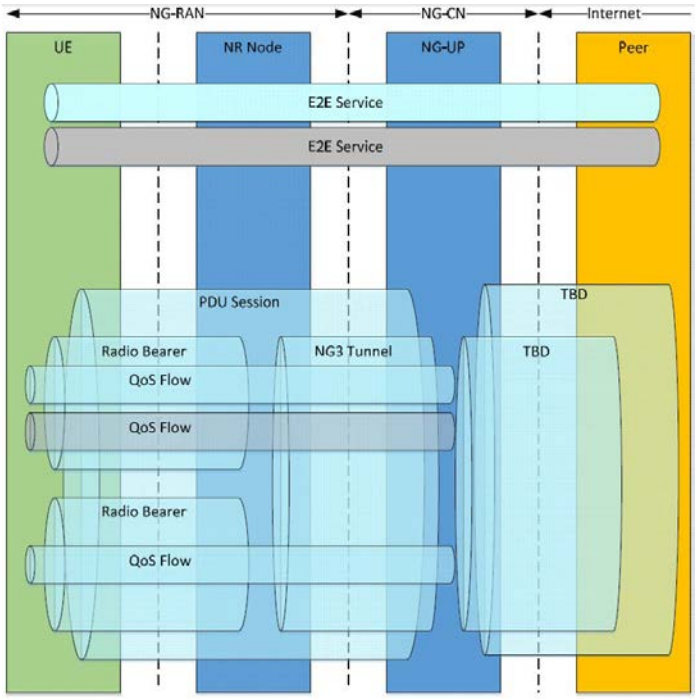


Fig. 3 QoS flow Architecture in 5G NG Network

III. 5G Enabling Technologies

The deployment and implementation of 5G wireless systems require re-engineering in the design of various existing network technologies and communication systems. Based on the service types previously stated together with the related requirements, key 5G RAN design requirements were outlined by 3GPP. Some of the primary approaches and technologies that have been developed to address the said are as follows:

A. Device to Device (D2D) Communication system

Standardised in 3GPP LTE Release 12 as LTE Device-to-Device proximity Services, Device-to-Device communication will enable mobile devices to discover the presence of the nearby devices within a defined range of up to 500 m and further directly communicate with them with minimal support from the network. Exploiting this device driven communication between devices improves overall throughput, spectrum utilisation as well as energy consumption, while still offering proximity based- and peer-to-peer applications and services [7] [8].

B. Massive MIMO

This is an evolving technology that is upgraded from current MIMO technology in which array of antenna are used containing few hundred antennas that are at the same time in one time and frequency slot serving many tens of user terminals [9]. This is particularly important in separating the indoor and outdoor scenarios during the 5G architecture design to ensure propagation loss can somehow be avoided [10].

Basically, it is a form of multiuser MIMO in which the

number of antennas ate the base station is much larger than the devices per signalling resource. Massive MIMO will offer robust, energy efficient, and secure and spectrum efficient Next generation networks [11] [12].

C. Ultra Dense Network

Heterogeneous networks will play a crucial role in achieving an ultra-dense networks. With the introduction of adhoc social networks and moving networks, the heterogeneous networks are becoming more dynamic. The network will consist of a large number of macrocells along with other low power nodes such as remote radio heads, relays and small cells. Though dynamic and dense heterogeneous networks will offer some new challenges in terms of backhauling, mobility and interference, new layer functionalities design would ensure performance maximization. Therefore, future smart devices will be designed in such a way they will learn and decide how to manage connectivity through the help of context information [13].

D. Machine Type Communication

Connecting machines, apart from people, is another important aspect of 5G. It's an emerging application in which either one or both of the bearers in a communication session involve machines. However, there is still a challenges in this type of communication as the number of devices that need to be connected are tremendously large, and there is a high demand for real time and remote control of mobile devices through the mobile network [6]. This will dictate a very low latency level essentially less than a millisecond, demanding a 20 times improvement in latency from the 4G network as we move to 5G network. This will form the basis of the Internet of Things with a broad range of applications including emergency services, automotive industry, medical field and public safety.

E. Millimetre Wave Solution

As a quest for more spectrum, this forms one of the key technology enablers for 5G. The conventional sub-3 GHz spectrum is getting increasingly congested, and the present radio access technologies are approaching the capacity limit of Shannon. 5G is exploring the mm- and cmWave bands to help solve the challenge [14]. However, the mmWave development is faced with three main impediments.

First, the penetration loss through the buildings is significantly higher at these spectrum bands, blocking the indoors users from the outdoor radio access technologies [15]. Second, the electromagnetic waves have the tendency to propagate in the direction of line of sight, causing vulnerability of radio links to being blocked by people and object movement. Lastly, compared to the conventional sub-3GHz bands, the path loss is relatively higher at these bands [16].

However, the enormous spectrum available in the mmWave band is a motivation as seamlessly ultra-broadband wireless pipes can be provided and completely revolutionize mobile communications. Further with the mmWave, very high beam forcing gains can be achieved as small sized antennas and their small separations can be utilised, enabling tens of antenna elements to be packed in just one square centromere [17].

F. Interface Management

In ensuring efficient utilisation of limited radio resource, reuse is a concept that is being utilised by various specifications for wireless and mobile communication systems [18]. For improved user throughput and traffic capacity, densification and reuse concepts will ensure enhancement in terms of efficient load sharing between local access networks and the macro cells [19]. Further, advanced sharing and joint scheduling could offer solutions in interference management.

G. Cloud-Based Radio Access Network

Evolving from distributed base station architecture in which the base station server is responsible for baseband processing, the cloud based radio access network will improve network capacity, enhance scalability and extend future 5G system coverage [20].

IV. 5G Radio Resource Allocation

In the previous networks, a radio resource is considered typically as part of the convention aspect of resource. It is usually characterised by frequency (bandwidth and carrier frequency), time (duration of the transmission), transmit power, and other parameters of the network such as modulation/coding schemes and the antenna configuration. In NR scheduling, both control and data are supported with the same numerology. Cross-slot and same slot scheduling is supported for both uplink and downlink transmission [6].

In a recent study on radio resource management for mmWave RAN that is being considered as a key enabler for 5G network, various challenges were raised [21]. Performance can be severely deteriorated by the high penetration loss of mmWave frequencies. Therefore, it was identified that maintaining a reliable connectivity would be difficulty particularly for services that are delay critical. The author noted that advanced interference management will be required in 5G system as the highly directional transmissions will make cross link interference characteristics become much different from systems of sub-6 GHz, introducing challenges such as flashlight effects [21]. During the movement of the users, link quality and wireless channels conditions can change significantly thus demanding multi connectivity support and fast RRM decisions. The study suggests the need to reengineer the conventional RRM functionalities for load balancing and connection management to overcome the challenges [21].

Auction algorithm method, message passing and stable matching mathematical models have been proposed in the development of radio resource allocation algorithms in the heterogeneous multi-tier 5G network [22]. The solution tool presented in the study are suggested for other enabling 5G system technologies in addressing the resource allocation problems [23] [24]. Challenges of resource allocation under the constraints of MU-MIMO and carrier aggregated 5G network were addressed in a recent study carried in Huawei

Labs [25]. A MU-MIMO aware CA scheduler is proposed in the study that assigns the resources available in frequency, time and MIMO dimensions under wireless system architecture and realistic traffic constraints. The comparative performance analysis indicated a user throughput of up to 190% compared to the baselines schemes, and a fairness of almost 1 using the Jains index value [25].

Efficient scheduling radio access technology in improving the quality of service has been investigated in [26], addressing the three scenario introduced in 5G network; dual connectivity (UE is served by NR and LTE at the same time), single connection (the UE is served only by one radio access technology), and fast radio access technology (in which the UE switches from one radio access technology to another) [16]. The Fig. 4 indicates instantaneous UE throughput comparing the fast radio access technology (FS-RAT) and the dual connectivity. The study concludes that 5G networks could merge the fast scheduling and the dual connectivity scheduling solutions in developing efficient scheduling algorithms for the 5G. The time between the consecutive evaluations of the RAT scheduling is also suggested to vary according to specific parameters as opposed to a fixed value [26].

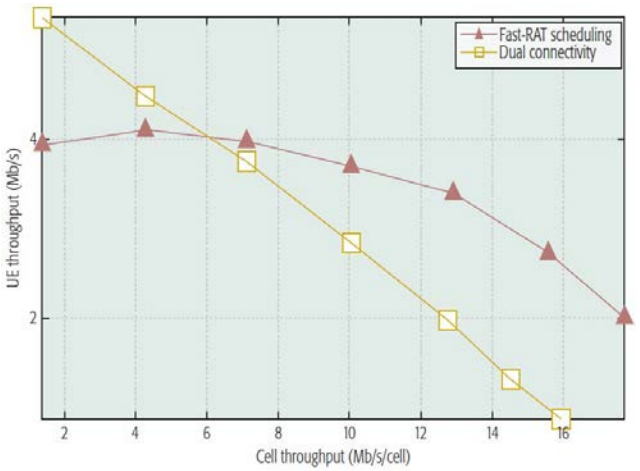


Fig. 4 UE Throughput Analysis for Dual connection and Fast RAT Scheduling [26]

The study in [27] tackles the problem of resource allocation for combined licensed and unlicensed CA MIMO systems through a Lagrange dual decomposition method for low complexity reasons. The allocation scheme proposed handles the coexistence matters within the unlicensed bands involving solution to a mixed integer nonlinear programming.

V. Allocation Optimisation in D2D Communication

In D2D enabled mobile network, radio resource management introduces certain challenges such as interference with D2D and cellular users, while reusing and underlaying the same radio resources with the users. To attain considerably high data rates in 5G future networks while ensuring the power and quality of service constraints, it is very important to intelligently select the specific operation mode for optimal scarce resource sharing.

Problems relating to D2D communications have been addressed in various studies, and different researchers have taken different perspectives of handling the issues, in effort of either maximising or minimising certain system parameters. Some of the problem aspects associated to the D2D resource management are power control, fairness, quality of service (QoS) and cooperative communication. In the various resource allocation schemes, the primary objectives are ensuring maximised system throughput, energy efficiency, and battery life, while minimising the overall system cost.

In [28], a D2D communication study is done focused on maximising the system throughput in the network. A distance dependent algorithm for mode selection was studied, ensuring QoS and power optimisation of the users were maintained. Dynamic power allocation, as opposed to fixed power control, was used to offer efficient resource utilization as well as better throughput gain.

Using shrinking reuse distances with frequency resources deployment, the authors [29] in presented a spectrally efficient radio resource management technique for device driven cellular network. The authors presented a D2D pair cluster formation and D2D pairs are grouped according to interference alignment similarity. This offers additional gains in throughput of the cellular system and a 33% extra rate increase compared a point to point type of communication. Frequency resources reuse over the clusters offers rate increase proportional to clusters [29].

Considering wireless channel uncertainties, authors in [30] investigated radio resource allocation challenges in a relay aided device driven network. Using stable matching, the study provided an iterative distributed solution to the problem. The study compares the proposed scheme against the underlay D2D communication scheme [30]. Stochastic geometry concept in a heterogeneous network is considered in enhancing the capacity of D2D mobile network while maintaining the constraint of outage probabilities of both cellular and D2D communication [31]. The authors present an algorithm that determines the set containing the critical points (believed to exhibit maximum capacity) and then subsequently obtain optimal solution from the points. The algorithm guarantees the QoS of service as the transmission reuse the licensed spectrum. The strategy manages resources by taking into account the co-tier and cross-tier interference [31]. Rate maximisation in D2D MIMO systems is studied in [32] while [33] discuss the sum rate maximisation in D2D enabled network. The paper applied the technique in which for better spatial reuse, short range D2D links are allowed to share the channels of other users in the same cell. On the pareto optimal boundary, there exists four points that comprise feasible SINR regions. The author argue that, out of the four points, one of the point comprises of the optimal solution to the non-convex optimisation problem. Using a greedy algorithm with low computational complexity,

authors suggest that the optimal solution can be obtained.

Particle swarm optimisation solutions for resource allocation in D2D communication have been suggested in [34] [35] [36]. Resource allocation and joint mode selection is studied to maximise the system throughput, while maintaining required data rates and minimum interference [34]. The main objective of the studies is to maximise the overall system throughput while managing the resource assignment, rate and QoS constraint. Power allocations and resource scheduling for the vehicular user equipment were studied in a recent study [37]. A dynamic radio resource allocation algorithm was developed based on Lyapunov optimisation theory, considering interference reduction in vehicle to vehicle (V2V) communications as well as reliability and latency requirements such communications.

Battery life as part of resource optimisation has been studied in the recent papers for the D2D communications [38] [39] [40] [41] [42] [43]. Energy efficiency have been studied with respect to various aspect in the studies. However, the general principle towards a longer battery life of a device is a factor of their transmit power. However, even though lower transmit power is desirable, the optimisation process will involve a trade off since higher data rates require higher SINR ratio and compromise transmit power minimisation [43].

VI. Conclusion

Adapted as an enabling technology in 5G Network, D2D communication enables communication between devices vicinity with low energy consumption and low latency, and potentially to offload cellular network from handling local network traffic. It is also anticipated to improve energy and spectral efficiency of cellular networks, while offering proximity based services such as file sharing and social networking. However, energy efficiency and interference management have been fundamental requirement in curbing the interference brought about by the device driven users under control, while still extending the lifetime of the user equipment battery.

The architecture of 5G RAN and other enabling technologies are presented. Radio resource management in 5G network is briefly introduced, and some of the issues studied in recent papers in regard to radio resource management in 5G presented. In the sequel of the paper, D2D communication is introduced. Various resource allocation schemes and strategies for the D2D communication in 5G network have been presented in this paper, and particularly in regard to constraints of the system throughput, energy efficiency, battery life, and the overall system cost.

While several techniques have been proposed in solving the optimisation problems in the device driven networks, the techniques deployed most in this study were based on heuristic algorithms, lagrangian formulation, graph theory, particle swarm optimisation, greedy algorithm, matching game, stable matching, game theory, stochastic search methods, among others.

The paper attempts to exploit an emerging field, and most areas are still unexploited. Improvements of battery life, energy efficiency improvements, optimal power allocations and energy harvesting are some of the interesting unexploited research domain. Segregation of indoor and outdoor antennas, as well as cell densification could also form future research areas.

References

[1] ITU-R Rec. ITU-R M. 2083-0, "IMT Vision — Framework and Overall Objectives of the Future," September 2015.

[2] Z. S. Bojkovic, M. R. Bakmaz and B. M. Bakmaz, "Research challenges for 5G cellular architecture," in 2015 12th International Conference on Telecommunication in Modern Satellite, Cable and Broadcasting Services (TELSIKS), Nis, Serbia , October, 2015.

[3] Akkarajitsakul, Phunchongharn, Hossain and Bhargava, "Mode selection for energy-efficient d2d communications in lte-advanced networks: A coalitional game approach," in Proceedings of IEEE International Conference on Communication Systems (ICCS), Singapore, Novermber, 2012.

[4] C.-C. Chien, H.-J. Su and H.-J. Li, "Device-to-device assisted downlink broadcast channel in cellular networks," in 2012 15th International Symposium on Wireless Personal Multimedia Communications (WPMC), Taipei, Taiwan, September, 2012.

[5] D. Lee, S.-I. Kim, J. Lee and J. Heo, "A transmission period selection scheme for device-to-device communications," in October, 2013, Jeju, South Korea , 2013 International Conference on ICT Convergence (ICTC).

[6] T. Olwal, K. Djouani and A. Kurien, "A Survey of Resource Management Toward 5G Radio Access Networks," IEEE Communications Surveys & Tutorials, vol. 18, no. 3, pp. 1656-1686, 2016.

[7] 3GPP, "5G; Study on New Radio (NR) access technology (3GPP TR 38.912 version 14.0.0 Release 14)," 3GPP, May, 2015.

[8] X. Li, R. Shankaran, M. Orgun, G. Fang and Y. Xu, "Resource Allocation for Underlay D2D Communication with Proportional Fairness," IEEE Transactions on Vehicular Technology, pp. 1-1, March, 2018.

[9] G. D. Swetha and G. R. Murthy, "D2D communication as an underlay to next generation cellular systems with resource management and interference avoidance," in 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET),, Chennai, India, March, 2017.

[10] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin and R. Zhang, "An Overview of Massive MIMO: Benefits and Challenges," IEEE Journal of Selected Topics in Signal Processing , vol. 8, no. 5, pp. 742-758, 2014.

[11] C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H. Aggoune, H. Haas, S. Fletcher and E. Hepsaydir, "Cellular architecture and key technologies for 5G wireless communication networks," IEEE Communications Magazine, vol. 52, no. 2, pp. 122-130, 2014.

[12] S. A. Busari, K. M. S. Huq, S. Mumtaz, L. Dai and J. Rodriguez, "Millimeter-Wave Massive MIMO Communication for Future Wireless Systems: A Survey," IEEE Communications Surveys & Tutorials, pp. 1-1, December, 2017.

[13] M. N. Boroujerdi, A. Abbasfar and M. Ghanbari, "Antenna assignment in Cell Free Massive MIMO systems," in 2017 Iranian Conference on Electrical Engineering (ICEE), Tehran, Iran, May, 2017.

[14] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," IEEE Communications Magazine, vol. 49, no. 6, 2011.

[15] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong and Jocelyn, "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!," IEEE Access, vol. 1, pp. 335-349, 2013.

[16] G. Marco, M. Marco, R. Sundeep and Z. Michele, "Multi-connectivity in 5G mmWave cellular networks," in Ad Hoc Networking Workshop

(Med-Hoc-Net), 2016 Mediterranean, Vilanova i la Geltru, Spain, JUne, 2016.

[17] N. Al-Falahy and O. Y. K. Alani, "Design considerations of ultra-dense 5G network in millimetre wave band," in Ubiquitous and Future Networks (ICUFN), 2017 Ninth International Conference on, Milan, Italy, July, 2017 .

[18] E. Hossain, M. Rasti, H. Tabassum and A. Abdelnasser, "Evolution toward 5G multi-tier cellular wireless networks: An interference management perspective," IEEE Wireless Communications, vol. 21, no. 3, pp. 118-127, 2014.

[19] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. Sukhavasi, C. Patel and S. Geirhofer, "Network densification: the dominant theme for wireless evolution into 5G," IEEE Communications Magazine, vol. 52, no. 2, pp. 82-89, 2014.

[20] M. Hadzialic, B. Dosenovic, M. Dzaferagic and J. Musovic, "Cloud-RAN: Innovative radio access network architecture," in 2013 55th International Symposium ELMAR, Zadar, Croatia, September, 2013.

[21] L. Yilin, P. Emmanouil, V. Nikola, L. Jian, X. Wen and C. Giuseppe, "Radio resource Managment Considerations for 5G Millimetre Wave Backhaul and Access networks," IEEE Communications Magazine, vol. 55, no. 6, pp. 86-92, 2017.

[22] R. Vannithamby and S. Talwar, "Distributed Resource Allocation in 5G Cellular Networks," in Towards 5G: Applications, Requirements and Candidate Technologies, West Sussex, Wiley Telecom, 2017, pp. 472-499.

[23] Y. Richard and L. Chengchao, "Wireless Network Virtualization: A Survey, Some Research Issues and Challenges," IEEE Communications Surveys & Tutorials, vol. 17, no. 1, pp. 358-380, 2014.

[24] H. Mesud, D. Branko, D. Merim and M. Jasmin, "Cloud-RAN: Innovative radio access network architecture," ELMAR, 2013 55th International Symposium, pp. 115-120, September 2013.

[25] A. Mustapha and F. Afef, "Downlink radio resource allocation with MU-MIMO and carrier aggregation in 5G networks," in 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Montreal, QC, Canada , October, 2017.

[26] M. Victor, E. Marten and R. C. Francisco, "Fast-RAT Scheduling in a 5G Multi-RAT Scenario," IEEE Communications Magazine, vol. 55, no. 6, pp. 79-85, 2017.

[27] Tsinos, Foukalas and Tsiftsis, "Resource Allocation for Licensed/Unlicensed Carrier Aggregation MIMO Systems," in IEEE Wireless Communications and Networking Conference, April 2016.

[28] S. Xiang, T. Peng, Z. Liu and W. Wang, "A distance-dependent mode selection algorithm in heterogeneous D2D and IMT-Advanced network," in Globecom Workshops (GC Wkshps), 2012 IEEE, Anaheim, CA, USA , December, 2012.

[29] Elkotby, Elsayed and Ismail, "Shrinking the reuse distance: Spectrally-efficient radio resource management in D2D-enabled cellular networks with Interference Alignment," in Wireless Days (WD), 2012 IFIP, Dublin, Ireland, November, 2012.

[30] M. Hasan and E. Hossain, "Distributed Resource Allocation for Relay-Aided Device-to-Device Communication Under Channel Uncertainties: A Stable Matching Approach," IEEE Transactions on Communications, vol. 63, no. 10, pp. 3882-3897, 2015.

[31] Z. Chen, H. Zhao, Y. Cao and T. Jiang, "Load balancing for D2D-based relay communications in heterogeneous network," in 13th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt) , Mumbai, India.

[32] A. Tölli, J. Kaleva and P. Komulainen, "Mode selection and transceiver design for rate maximization in underlay D2D MIMO systems," in IEEE International Conference on Communications (ICC), London, UK, JUne, 2015.

[33] Cheng, Lin and Gu, "Power and channel allocation for device-to-device enabled," Journal of Computational Information Systems, vol. 10, no. 2, pp. 463-472, 2014.

[34] S. Sun and Y. Shin, "Resource allocation for D2D communication using Particle Swarm Optimization in LTE networks," in Information and Communication Technology Convergence (ICTC), 2014 International Conference on, Busan, South Korea , 2014.

[35] L. Su, Y. Ji, P. Wang and F. Liu, "Resource allocation using particle swarm optimization for D2D communication underlay of cellular networks," in IEEE Wireless Communications and Networking Conference (WCNC), Shanghai, China, 2013.

[36] R. Tang, J. Dong, Z. Zhu, J. Liu, J. Zhao and H. Qu, "Resource Allocation for Underlaid Device-to-Device Communication by Incorporating Both Channel Assignment and Power Control," in Fifth International Conference on Communication Systems and Network Technologies (CSNT), , Gwalior, India , April, 2015.

[37] N. Yu, J. Mei, L. Zhao, K. Zheng and H. Zhao, "Radio resource allocation for D2D-based V2V communications with Lyapunov optimization," in IEEE/CIC International Conference on Communications in China (ICCC), Qingdao, China, China , October, 2017.

[38] Y. Wu, J. Wang, L. Qian and R. Schober, "Optimal Power Control for Energy Efficient D2D Communication and Its Distributed Implementation," IEEE Communications Letters, vol. 19, no. 5, pp. 815-818, 2015.

[39] G. Fodor and N. Reider, "A Distributed Power Control Scheme for Cellular Network Assisted D2D Communications," in 2011 IEEE Global Telecommunications Conference (GLOBECOM 2011), Kathmandu, Nepal , 2011.

[40] X. Xiao, X. Tao and J. Lu, "A QoS-Aware Power Optimization Scheme in OFDMA Systems with Integrated Device-to-Device (D2D) Communications," in IEEE Vehicular Technology Conference (VTC Fall), San Francisco, CA, USA , September, 2011.

[41] Y. Luo, P. Hong, R. Su and K. Xue, "Resource Allocation for Energy Harvesting-Powered D2D Communication Underlaying Cellular Networks," IEEE Transactions on Vehicular Technology, vol. 66, no. 11, pp. 10486-10498, 2017.

[42] F. Wang, C. Xu, L. Song, Q. Zhao, X. Wang and Z. Han, "Energy-aware resource allocation for device-to-device underlay communication," in IEEE International Conference on Communications (ICC), Budapest, Hungary , 2013.

[43] A. Abdallah, M. M. Mansour and A. Chehab, "Power Control and Channel Allocation for D2D Underlaid Cellular Networks," IEEE Transactions on Communications, pp. 1-1, 2018.

A Review of Control Strategies for Microgrid with PV-Wind Hybrid Generation Systems

Linus A. Alwal, Peter K. Kihato and Stanley I. Kamau

E-mail: [1]laloo@jkuat.ac.ke [2]pkihato@jkuat.ac.ke [3]skamau@eng.jkuat.ac.ke

Abstract— The concept of microgrid has been attracting a lot of interest among researchers and investors over the last decade. This is because it offers a promising technology towards the utilization of distributed renewable energy resources, notably Photovoltaic (PV) and Wind hybrid generation systems in islanded or grid-connected mode. The main problem currently facing utilization of PV-Wind based distributed generation systems is that of consolidating the power and controlling its voltage and frequency levels on the right scale for either off-grid or on-grid utilization. This is due to their intermittency in power outputs and geographical fragmentations. Thus there is need to coalesce such power units into microgrids with reliable storage systems which can be better managed using a microgrid control system.

This study aims at reviewing the advances made towards the utilization of microgrid control systems incorporating PV-Wind generation units with emphasis on the applicable microgrid control strategies and principles. Based on the review conducted, it was noted that every control strategy has its merits, demerits, level of efficiency and cost. Further, to address the control challenges of the PV-Wind hybrid generation system, a Multilevel Microgrid Control System (MMCS) utilizing appropriate artificial intelligence has been proposed. The research directions in this field are geared towards the use of artificial intelligence to optimize performance of the system.

Keywords— Control Strategies, Hybrid, Microgrid, PV-Wind

I. INTRODUCTION

Nowadays, there are increasing needs to shift focus on power generation from the environmentally unfriendly conventional sources such as diesel or coal plants to the clean and renewable sources such as photovoltaic (PV) and wind [1]-[6]. However, the main bottleneck in the utilization of these renewable resources has been on the proper control and management of their power outputs as well as robust protection mechanisms. This is because owing to the intermittent nature of PV and wind resource, there is no guarantee that the needs of a given load or power system will be efficiently served [7], [8]. The control and protection problem is even more complicated when a hybrid of such sources is adopted. One solution to this has been on pooling together the resources into a microgrid. In its basic form, a microgrid may be viewed like a small local power system whose main objective is to serve or electrify remote communities [4], [5]. Microgrids can be operated in two modes: grid-connected mode and islanded mode [9]-[13]. For efficient performance of a microgrid, a reliable microgrid control system with suitable control algorithms is required [14]-[18]. The aim of this study is to provide an overview of such control strategies which are applicable to a hybrid microgrid system. The control strategies and architectures reviewed include hierarchical control schemes, the conventional and modified droop control methods, model predictive control, nonlinear programming algorithms, artificial intelligence control methods, and commercial microgrid systems. Also, control design guidelines particularly those that contain renewable energy sources such as PV and wind in hybrid configuration are discussed. Further, based on the observed trends and gaps, a Multilevel Microgrid Control System (MMCS) utilizing suitable artificial intelligence has been proposed for PV-Wind hybrid generation microgrid system. The future research directions, challenges and conclusion are captured towards the end of the study. The subsequent sections of the paper are organized as follows: Section II presents Hybrid PV-Wind Microgrid System while Section III captures Existing Microgrid Control Strategies. In Section IV, Microgrid Challenges and Proposed Control Framework are presented. Finally, the review is wrapped up in Section V with Conclusions and Future Research.

II. HYBRID PV-WIND MICROGRID SYSTEM

A. *Microgrid Control System (MCS)*

A Microgrid Control System (MCS) is a supervisory based control architecture which enables efficient and cost effective utilization and integration of Distributed Energy Resources (DERs), loads and energy storage in a localized framework [19]-[22]. Fig. 1 shows a microgrid control system with solar PV and wind as the principal microsources [5]. A microgrid control system is required to regulate and optimize the utilization of power outputs from these sources which may be coupled directly to the DC bus at the point of common coupling (PCC) and subsequently to the AC bus via an inverter. The loads, classified as low priority, essential AC and dump, are connected to the load bus. The main roles of a microgrid control system are as follows [2]-[6]: First, is control of voltage and frequency for both grid-connected and islanded modes. Second, it provides efficient load sharing, co-ordination of DERs, and

Linus A. Alwal, Department of Electrical and Electronic Engineering, Jomo Kenyatta University of Agriculture & Technology (JKUAT), Kenya, (phone: +254729747653; e-mail: laloo@jkuat.ac.ke);

Peter K. Kihato, Department of Electrical and Electronic Engineering, JKUAT, Kenya (e-mail: pkihato@jkuat.ac.ke);

Stanley I. Kamau, Department of Electrical and Electronic Engineering, JKUAT, Kenya (e-mail: skamau@eng.jkuat.ac.ke).

microgrid re-synchronization with the main grid. Third, it controls power flow between microgrid and main grid and also optimizes its operating cost. Finally, it appropriately handles transients and re-establishes prescribed conditions when transitioning between the modes.

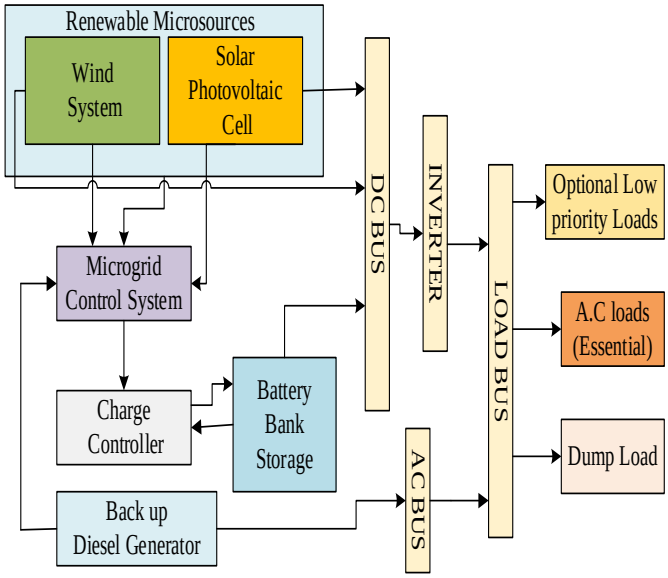


Fig.1 Block diagram of Microgrid Control System [5]

Two broad categories of microgrid on the basis of the type of power handled are AC microgrid and DC microgrid, which gives rise to AC and DC microgrid control techniques [2], [10], [11], [18] and [19]. On the other hand, in terms of the operation principle, microgrids can operate in two modes: grid-connected mode and islanded (off-grid) mode [15]-[18]. Based on the grid connection status of a microgrid, it can be categorized into two. First, as permanently islanded microgrid in which stand-alone networks serve as the sources for all of the locally generated power which in turn must be consumed by the loads in the isolated network. Second, is the grid-tied microgrid which is not only capable of generating power within its distribution networks but also import power from a utility source [18]. The former networks are typically found in remote areas or island communities where high transmission costs and losses render connection to the grid or bulk importation of fuel uneconomical. Further, renewable resources (wind, solar, hydro) should be available to make optimization of generation resources very desirable [1], [2] and [3].

B. *Photovoltaic and Wind Energy System*

The modeling of a Hybrid PV-wind microgrid system shown in Fig. 1 can be done by considering the PV and wind energy systems separately. However, there is interdependency in the sizing, control setting and operation strategies [5]. Consequently, the choice among design possibilities of a given hybrid PV-wind system becomes quite challenging. The important aspects of these systems are the power outputs as well as their equivalent circuit models.

The power generated from a PV system with respect to solar radiation (SR), denoted by P_{PV}, is given by (1).

$$P_{PV} = P_{STC}\frac{G_{ING}}{G_{STC}}\left(1 + k\left(T_C - T_\gamma\right)\right) \quad (1)$$

where P_{PV} is the power generated from a PV system, G_{ING} is the incident irradiance, P_{STC} and G_{STC} is maximum power and irradiance at standard test condition respectively, k is the temperature coefficient, T_C is the module temperature and T_γ is the reference temperature.

The power generated from a wind turbine (WT) unit, P_{WT} can be specified as a cubic function with respect to the wind speed at the monitoring station as in (2).

$$P_{WT} = \frac{1}{2}\rho\pi R^2 v^3 C_p \quad (2)$$

where P_{WT} is the power generated from a wind turbine (WT) system, R is the radius of the paddle of wind turbine, ρ is the air density of the monitoring area, v is the wind speed of power generation area and C_p is conversion efficiency of the wind power.

III. MICROGRID CONTROL METHODS

A. *Hierarchical Control Schemes*

The control objectives of AC and DC microgrids can be considered in terms of a hierarchical or multilevel control structure, which can be classified into three levels as: primary, secondary and tertiary control levels [2], [3]. This leads to a multilevel control structure shown in Fig. 2.

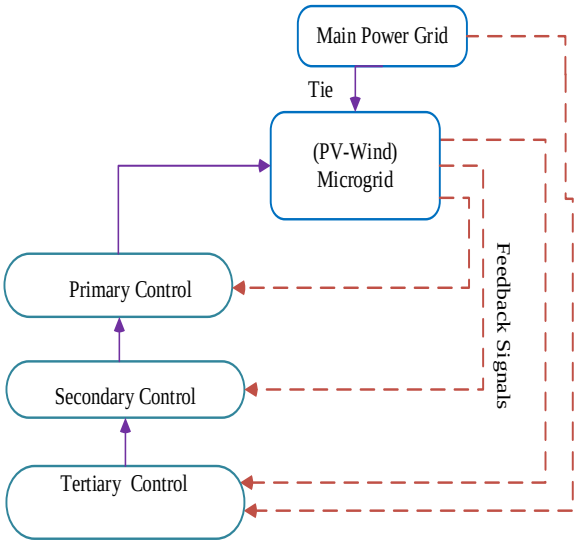


Fig.2 Hierarchical control levels of a microgrid [2]

The *primary control level* operates at the fastest timescale and encompasses the fundamental control hardware for executing the internal voltage and current control loops of the distributed PV-wind microsources. Consequently, it deals with local measurements and does not need a communication system. It is responsible for ensuring that the voltage and frequency are stabilized within the microgrid or when switching mode. It also

curtails the occurrence of circulating currents and provides a plug-and-play capability of the DERs. The need to provide independent active and reactive power sharing controls for the DERs is emphasized especially in the presence of both linear and nonlinear loads [2]. Principally, the most common strategies for active power and reactive power regulation are droop-based controls (voltage-reactive power and frequency-active power droop controls) [1]-[4].

The *secondary control* is utilized to compensate for the variations in voltage and frequency due to actions by the primary controls. However, this task becomes particularly challenging in stand-alone PV-Wind microgrids due to the presence of highly-variable sources [4], [5]. It can be realized using two main approaches: centralized and decentralized architectures, essentially defined depending on the position of the Microgrid Central Controller (MGCC) [4], [8]-[9]. Centralized control is characterized by complex central processing elements which are situated far away from the microsources. It uses a communication system such as a Microgrid Supervisory Control And Data Acquisition (MicroSCADA) which involves a central system, sensors and control devices [9], [12]. It is recommended for certain small microgrids with shared goals and thus should cooperate. However, with growth in microsources data, the centralized approach tends to be slower and may easily malfunction. Apart from creating a single point of failure to the microgrid system, it also hinders up-scaling and may be uneconomical. This is where the decentralized control, which favors the utilization of the proposed PV-wind microgrid, comes in.
Decentralized control is preferred for microgrids of different vendors and which should make independent decisions pertaining to their operating conditions. In other words, decentralized microgrid control system incorporates an intelligent control for each microsource. It reduces the network complexity, allows easy scalability, improves power supply reliability, more economical and offers bi-directional power flow [9]. However, a disadvantage of this approach is the load-dependent frequency and amplitude deviations due to the utilization of the droop control method in primary control to manipulate the active and reactive power [8].

Finally, at the highest level and corresponding to the slowest timescale, is the *tertiary control* which manages the bidirectional power flow between the microgrid and the main grid thereby facilitating an economical and optimal operation of the hierarchical control system. For instance, in a case where the power generated in the microgrid surpasses the local power demand, which coincides with maximum absorption of power from PV and wind renewable resources, then the excess power is transmitted directly to the high-inertia DC system or to the main AC grid through an inverter [2].

As depicted in Fig. 3, a number of control techniques have been proposed in literature for the inner-loop control and primary control, deployed according to the characteristics of the microgrid [19]-[25]. The aim is to improve the power quality, disturbance rejection and voltage or current tracking of the inverter output [13], [26]. Some of the techniques applicable to PV-wind microgrid are discussed next.

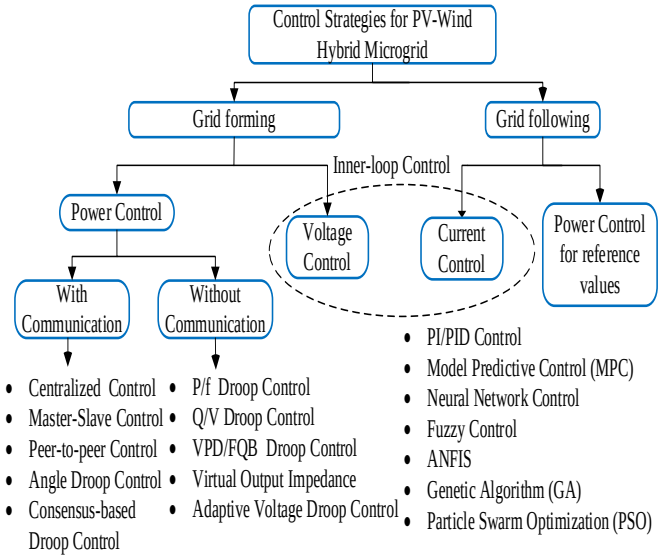


Fig.3 Inner-loop control and primary control techniques [13].

B. Droop Control Methods

Conventional and modified droop control methods have been widely utilized in microgrid systems particularly for inner control loops/primary control [1]-[4], [8]-[11], [13]-[16], [24]. The design of droop controllers (conventional voltage and frequency droop controllers) can be conducted using appropriate and separate small-signal models [2]. Important to note is the prevailing basic trade-off between the time constant of the control system and the frequency regulation. In microgrids, this is mostly implemented either in active/reactive power (PQ) mode or voltage control mode.

i. Active/reactive power (PQ) Control Mode: The microsource active and reactive power supply is regulated based on predetermined reference points as shown in Fig. 4. This is achieved using a current-controlled voltage source inverter (CCVSI) [2].

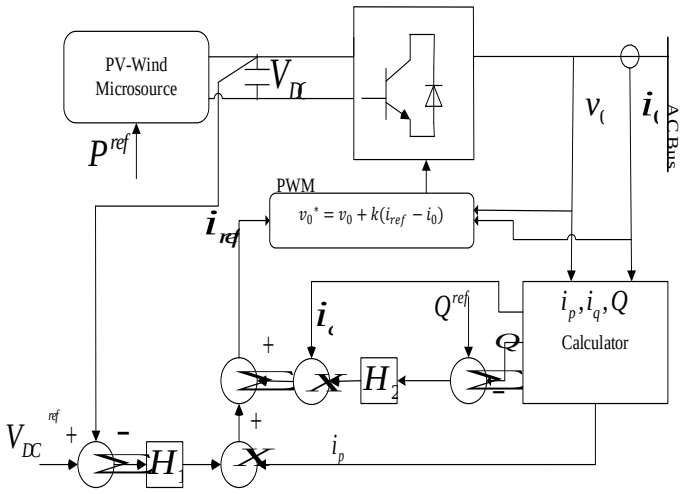


Fig.4 PQ control mode with active and reactive power references [2].

ii. Voltage Control Mode (VCM): The DER functions as a voltage-controlled voltage source inverter (VCVSI), implying that the reference voltage v_0^* is obtained through

droop characteristics using the primary control as shown in Fig. 5. Such techniques are acquiring more attention as they are capable of emulating the behavior of a synchronous generator [16].

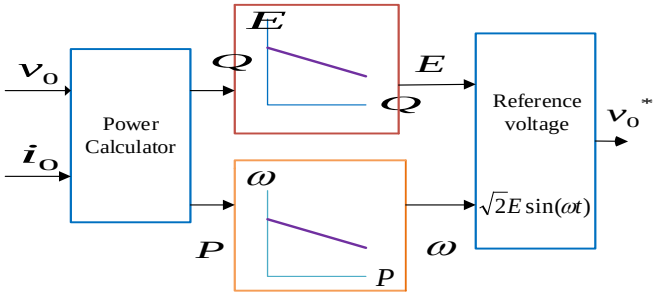


Fig.5 Determination of reference voltage for VCM [2]

The VCM nested voltage and current control loops is shown in Fig. 6 [2]. It is seen that such a control strategy injects the current signal as a feedforward term through transfer function. If this model were to be applied in small-scale islanded systems with PV-wind sources, the major concern would be on how to improve the power quality. This scenario is made even more challenging in presence of non-linear and single phase loads or the low inertia presented by the microgrid.

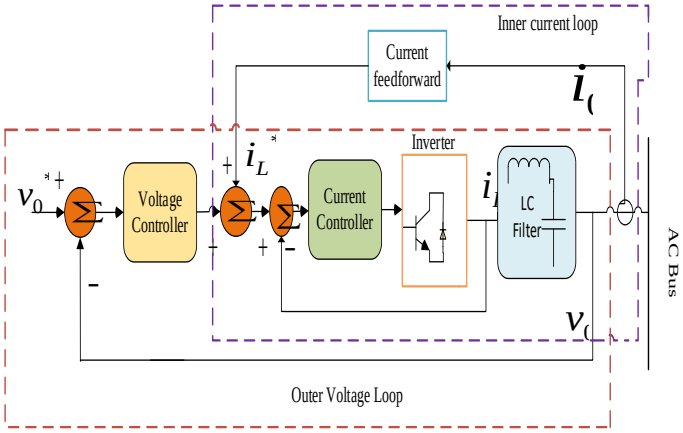


Fig.6 Voltage and current control loops in VCM [2]

One direct way to improve the power quality suggested in [2] and [13] is to modify the control structure by adopting a low pass filter both in the feedforward inner current loop as well as the feedback outer voltage loops which aids in the removal of harmonic contents. Contrary to the active load sharing technique, the conventional droop method can be implemented without any communication links, hence is more reliable. However, some of its shortcomings include [2], [13]-[16]:

i. The approach handles only one control variable for each droop characteristic thus it is impossible to simultaneously achieve multiple control objectives.
ii. Development of conventional droop method assumes that the effective impedance between the VCVSI and the AC bus is highly inductive. However, this assumption does not hold in microgrid applications where the low-voltage transmission lines are majorly resistive.
iii. In a microgrid, frequency is a global quantity whereas the voltage is not. Consequently, for critical loads, the reactive power control may adversely affect the voltage regulation.
iv. Further, for non-linear loads, the conventional droop method cannot distinguish the load current harmonics from the circulating current. This leads to distortion of the DER output voltage by the current harmonics.

The following techniques have been proposed to address the draw backs of the conventional droop method by minimizing the total harmonic distortion (THD) [2], [13]-[16].

i) *Adjustable load sharing method*-in which the time constant of the active and reactive controllers can be altered without affecting the DER voltage and frequency.
ii) *VPD/FQB Droop method*-The voltage active power droop and frequency-reactive power boost (VPD/FQB) characteristics approach is proposed for tackling the challenges of predominantly resistive low-voltage microgrid lines. The VPD/FQB technique can also be altered to vary the controller time constant without changing voltage and frequency. However, it is highly dependent on system parameters thereby restricting its application. Moreover, it can malfunction when subjected to non-linear loads and thus cannot assure sustained voltage regulation. Fig. 7 a) and b) show voltage-active power droop characteristic and frequency-reactive power boost characteristic respectively [2].

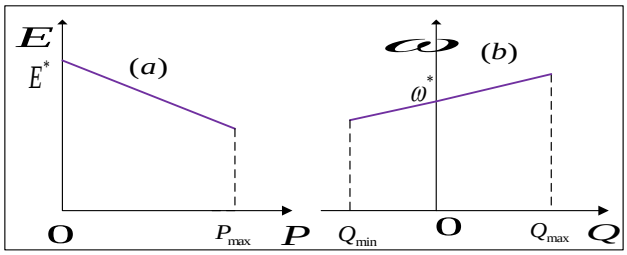


Fig.7 Low-voltage microgrids droop/boost characteristics [2].

iii) *Virtual Frame Transformation Method*-The droop method with virtual power frame transformation utilizes an orthogonal linear transformation matrix to transfer the active/reactive powers to a new reference frame. This results to the powers being independent of the effective line impedance [2].
iv) *Virtual Output Impedance*-which utilizes an intermediate control loop to vary the output impedance of the VCVSIs. This method mitigates against the dependency of the droop techniques on the system parameters and operates satisfactorily in the presence of non-linear loads. However, the method does not guarantee regulation of voltage. Also, adjusting the closed-loop time constant may cause undesired deviation in the DG voltage and frequency.
v) Some other techniques proposed include *adaptive voltage droop control* and *non-linear load sharing* [2].

C. Model Predictive Control and Nonlinear Programming

The Model Predictive Control (MPC) is presented in [4], [6], [19] and [20]. The MPC method uses a prediction model which is based on existing system knowledge and future predictions to determine the control variables in real-time (online) for every sampling period. From this, the minimal value of control variable in the predictive function is selected for the next

sampling period (time step) [6], [19]. Thus at every discrete time step, an optimization model which covers finite horizon is deployed to generate a control action sequence, from which often the first action is utilized. Using the renewed system state and future information, the system then moves to the next time step and the above computation is repeated [20]. Fig. 8 shows the principle of MPC which can be viewed as a closed-loop system due to continuous modulation of control variables to compensate for prediction inaccuracies.

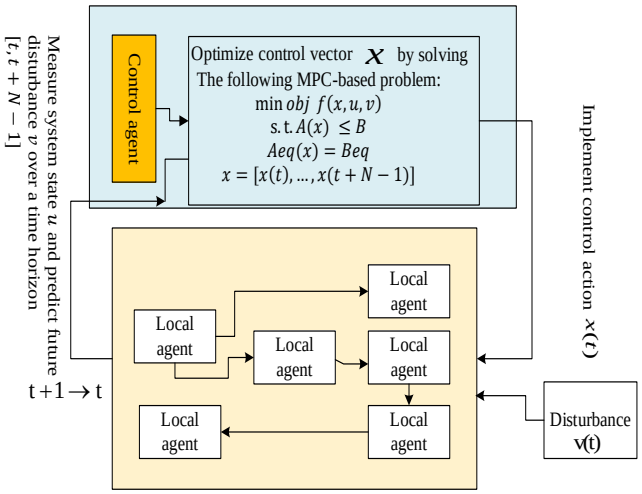


Fig.8 Principle of MPC [20]

The main components of the MPC scheme include:

i) a control agent which optimizes control actions over a finite horizon based on a system-wide objective function;
ii) multiple local agents to execute control actions;
iii) possible disturbances that may influence the local implementation of control actions.

The MPC method can be realized in centralized or decentralized schemes. It has been widely used as an inverter control strategy [6] and to solve the problem of real-time economic dispatch and power exchange between the microgrid and the energy storage (ES) or utility system [19], [20]. In the case of a microgrid economic dispatch, microgrid Energy Management System (EMS) acts as the central control agent where the local agents include micro turbines, energy storage system and scalable load. The central control agent aims at minimizing the total operation cost by generating suitable reference points for all controllable units and also the optimal power exchange with the utility grid. These agents together maintain a global power balance by consistently responding to the coordinated energy management commands from the higher level EMS. Wind speed, load changes and power prices constitute uncertain disturbances and their predications are constantly updated by EMS.

One of the existing MPC methods for microgrids with distributed ES involves strategies considering only one ES system or aggregated multiple ES systems. This doesn't account for the power flows between various ES systems. The other category involves strategies based on non-convex optimization. For example, MPC based on non-linear programming (NLP) in unbalanced microgrids [4], [19]. However, if the problem is non-convex scalability is limited and the available solvers are capable of only providing locally optimal solutions. Recursive dynamic programming (RDP), which produces a globally optimal solution, may be an alternative for the ES system optimal power flow problem. However, its numerical complexity increases with the number of ES systems [19].

In [19], a new convex MPC strategy is proposed for solving the dynamic optimal power flow problem between battery energy storage systems (BESS) distributed in an AC microgrid with intermittent PV generation. The problem formulation is based on a linear $d-q$ reference frame VCM and linearized power flow approximations. The results indicated that the proposed control strategy approaches the performance of a strategy based on non-convex optimization, while minimizing the required computation time by a factor of 1000. Since the problem was solved as a convex quadratically constrained quadratic program (QCQP), for which fast and robust solvers exist, it is recommended for a real-time receding horizon MPC implementation [19]. The method takes into consideration line losses, voltage constraints and converter current constraints during the optimization. The implementation of MPC for solving nonlinear optimization problems at each time-step is called nonlinear MPC (NMPC) and is presented in [4].

In [9] utilization of *Multi-Agent System (MAS) based Control* strategy is presented. This is deployed in hierarchical control architecture when several microgrids organized into clusters are to be controlled. It is made up of four agents; control agent, DER agent, user agent and database agent. This approach offers flexibility in microgrid control in grid-connected or stand-alone mode as well as providing a seamless transition between the two modes.

D. Artificial Intelligence Control Methods

When PV-wind hybrid microgrid systems are to be implemented in remote areas which are ideal for stand-alone operation, it becomes challenging to acquire long-term weather data such as solar radiation and wind speed to aid in sizing. Consequently, artificial intelligence techniques such as artificial neural networks (ANN), fuzzy logic [FL], genetic algorithms (GA) and particle swarm optimization (PSO) become very useful in sizing stand-alone systems as compared with the conventional sizing method which is highly dependent on long-term weather data [5], [7], [21]-[23].

a) Genetic algorithm method

Genetic algorithm (GA) method is a search technique and is suitable for a complex problem such as a PV-wind hybrid system when other techniques do not offer acceptable solutions. In this case, the weather conditions are varying hourly and daily and thus will also be different for different seasons in a year. The deployment of GA method aids in obtaining the optimum number of facilities to use based on the hourly average metrological and load data collected, say over a few years, for simulation purposes [5].

b) Artificial Neural Networks (ANN)

Artificial Neural Network (ANN) has a self-learning feature which enables the algorithm to be easily developed for different operating conditions and grid disturbances [13]. However, it lacks performance in off-line training technique. In [21], the Lagrange programming neural network (LPNN) is used in a hybrid microgrid to achieve optimal scheduling and management. The objective function incorporated power generation cost, operation and maintenance cost, emission-control cost and fuel cost. The LPNN minimized the cost function and maximized the energy generated by the wind turbines (WTs) and photovoltaic cells (PVs). The energy stored into and supplied from the storage system was optimized too. Further, day-ahead prediction of renewable resources and load demand was obtained using the radial basis function neural network (RBFNN). The results of LPNN were compared with those of the basic PSO. It was observed that the LPNN was better than the PSO method. For instance, when the LPNN cost is $274.64, the PSO cost is $476.84, which is 1.74 times higher. This indicates clearly that the PSO is weaker in dealing with constrained problem [21].

c) Particle Swarm Optimization

The particle swarm optimization (PSO) method is reported in [5], [21] - [23] and can be used in optimal sizing of the hybrid energy systems. In [22], the PSO method is used to solve a day-ahead microgrid dispatch problem taking into account uncertainties which are associated with energy production by the PV and wind sources. The stated uncertainties were modeled by incorporating a robust approach in PSO and the method tested on 21-bus microgrid (MG). From the test results, it was observed that the robust optimization approach (PSO) was more advantageous (17%) compared to the deterministic optimization. Although the execution time of PSO (2360 seconds) is comparatively high, it was acceptable for the day-ahead decision time period.

The success of PSO scheme is also reported in [23], where it is used in the design and optimization of a hybrid microgrid system (HMGS) using the solar and wind meteorological data. This study applied PSO in sizing of wind turbines (WT), photovoltaic (PV) module, battery energy storage system (BESS) and diesel generator. Based on simulation results and the sensitivity analysis conducted, it was observed that the PSO technique achieved the best size and configuration of PV-wind based HMGS.

d) Fuzzy Based Controllers

The Fuzzy Control Methods possess the ability to manage the non-linear behavior of complex control structures since they take advantage of heuristics and expert knowledge of the process under control. It is also insensitive to variations in system parameters. However, the strategy is relatively slow. In [24], reference is made to a Fuzzy Logic Controller (FLC) used in the non-linear DG interface for voltage regulation, control of real and reactive power as well as an Adaptive Linear Neuron (ADALINE) utilized to eliminate harmonics and unbalance compensation. The FLC is also implemented together with the conventional Proportional-Integral (PI) controller to regulate voltage and frequency in AC microgrid. In addition, a new intelligent droop control deploying adaptive neuro-fuzzy inference system (ANFIS) is discussed in [24]. This is used to provide a solution for intelligent model-free based generalized droop control (GDC) and achieves desired voltage and frequency regulation in an islanded microgrid.

E. Tools For Optimization of Microgrid Systems

Several software tools have been developed to aid in design, optimization and performance evaluation of microgrid systems [5], [25]-[27]. In [5], a detailed coverage of the software tools for optimization of system with predefined configurations is presented. The software tools include: SOMES, HOMER, HYBRID 2, INSEL, SOLSIM, WATSUN-PV, PVSYST, PV-DESIGN PRO, RAPSIM, PHOTO, RAPSYS, RETScreen and ARES. From these, only the first two (SOMES and HOMER) are recommended for PV-wind hybrid system since both have the ability to provide optimal design of the hybrid system. For instance, HOMER can be used to model both conventional and renewable energy technologies: including PV, wind turbine, diesel generator, fuel cell, utility grid, micro turbine, battery bank and hydrogen storage. It conducts simulation for all of the possible system configurations, including whether off-grid or on-grid and determines a feasible one. Next, HOMER estimates the installation and operation cost of the system and displays a list of configurations arranged according to their life cycle cost [5], [7].

In [25], the Grid IQ Microgrid Control System (MCS) developed by General Electric (GE) is presented. This is a commercial supervisory control architecture based on U90^{PLUS} Generation Optimizer that provides optimization solution for permanently islanded or grid-connected microgrids. The software is capable of forecasting load, renewables and electricity price; carry out microgrid generator unit commitment integration, energy storage integration, intelligent local controllers as well a suite of security features.

In [26] the “Microgrid Plus” control system based on MGC600 series of controllers, developed by ABB Australia Pty. Ltd is introduced. This is a commercial modular and networked power flow and energy storage control system which is suitable for stable integration of intermittent renewable generation (e.g. PV and wind) into microgrids. The system is designed for use in isolated microgrids such as in remote communities as well as in predominantly grid-connected microgrids such as for industrial complexes.

In [27], the state of commercial microgrid controllers is captured, exposing the strengths and weaknesses of adopting such microgrid control strategies. Currently, most suppliers offer a microgrid control solution instead of a set of microgrid control products that can be accessed off-the-shelf and utilized directly in a project by the system owner’s engineers. This denies the system investors some level of flexibility, scalability and accessibility in implementing controls. In addition, although the proprietary approaches offer a high level of security and reliability, there are challenges regarding the long-run sustainability and maintainability of the proprietary aspects of the control system. Any change in the microgrid system characteristics raises maintenance costs due to continued vendor involvement. For example, there may be a need to incorporate a new generation unit such as a PV system or wind turbine or load (due to increased demand) or a new component (due to technological changes). This hinders scalability to the microgrid system.

F. Other Control Strategies and Emerging Issues

ANFIS based Generalized Droop Control (GDC) is referred to in [28]. It is used to overcome the drawbacks of GDC-based frequency and voltage control in microgrids utilizing more than one DG (e.g. PV and WT). The ANFIS-based GDC deploys the training ability of the ANN to the FL to create a new hybrid technique, termed ANFIS and is trained using input-output (I/O) data saved from the GDC approach. This strategy is applicable to a wide range of MGs and does not require knowledge about the MG structure as well as the line parameters [28].

Microgrid islanding control can be achieved through the peer-to-peer strategy or master-slave strategy [29]. The peer-to-peer strategy allows for plug-and-play feature which means that microsources may be added to the microgrid without any adjustments to the existing control and protection set up. This may be implemented using droop control. In the master-slave strategy, there is a master controller while the others are slave controllers which receive instructions from the master controller via a communication link. In islanded mode, the master controller in the microgrid maintains the system frequency, and also power balance separately using suitable algorithm. The future research direction is suggested to focus on improving the operation, control and optimization of different DG systems including renewable intermittent (PV or WT), controllable, conventional and converter mode DG [29].

IV. MICROGRID CHALLENGES AND PROPOSED CONTROL FRAMEWORK

A. Microgrid Challenges

The development and utilization of microgrid systems (hybrid or otherwise) cannot be undertaken without technical, design, integration, protection, social, policy and sustainability challenges [1], [4], [28], [30]-[33]. In general, the generic challenges of a microgrid system have a huge stake in hybrid microgrid systems too. In [28], the technical challenges facing microgrid are summarized. Some of the technical challenges regarding hybrid microgrids include [4] and [28]:

a) Unstable operation when faults and various disturbances occur in the absence of storage elements.
b) Sophisticated control strategies are required to ensure stability and power quality especially in the islanded mode of operation.
c) Lack of seamless transitions from grid-connected to isolated mode of operation may lead to large mismatches between generation and load. This presents serious challenge to voltage and frequency control.
d) Difficulties in coordinated control in situations involving a large number of intermittent microsources.
e) Limitations in communication lead to adoption of distributed intelligent strategies.
f) In low voltage networks, the resistance to reactance ratio is high, which challenges decoupled and independent control of active power, P and reactive power, Q.

The challenges of microgrids in remote communities is presented in [30]. Here, the STEEP model, which stands for social (S), technical (T), economic (E), environmental (E) and policy (P) is proposed and used to identify, classify and achieve sustainable microgrids. The microgrid failure factors, sustainable microgrid factors and sustainable planning network are presented in detail. The issues regarding the modeling and simulation of hybrid microgrid systems are given in [31]. Most important in hybrid microgrids is the challenge of establishing the operation modes or regimes for the entire energy conversion system. Also, is the linearization of the mathematical model in the electrical machine modeling of wind energy system to reduce their order. The challenges associated with integration of Renewable Energy Resources, recognizing that solar and wind power are intermittent and non-dispatchable, are detailed in [32]. These include high variability of wind power, changes in wind generator frequency, time lag between solar generation peak and late afternoon demand peak, rapid solar output variation due to passing clouds, power quality reduction and limited forecasting abilities. Finally, challenges regarding design of microgrid protection systems are highlighted in [1] and [4]. They are due to: bidirectional power flow, network topology change (mesh or ring network), and converter interfacing (due to extremely low inertia of interfaced inverters).

B. Proposed Control Framework

From the literature reviewed, it is clear that every control strategy has its merits and demerits. Consequently, every approach should be assessed based on the needs of a microgrid set up for which it was designed. The first step in choosing a control strategy is to determine whether the microgrid will be grid connected or islanded or making transitions between the two. If on-grid, the microgrid control strategy is determined by the utility whereas in islanded mode of operation, the control strategy is determined by a set of factors. These include type of microsources and net capacity, capacity of available energy storage, type of load and ownership.

Based on shortcomings of existing microgrid control methods and considering the above factors, the strategy proposed in this study for control of a microgrid with PV-wind hybrid generation systems is a scalable Multilevel Microgrid Control System (MMCS) shown in Fig. 9. This should use MPC-ANFIS controller supported with either PSO or GA. The PSO or GA should be used to obtain optimal size of PV and wind turbines of the hybrid microgrid system based on the available meteorological data. The MPC should be used to modify the PQ or droop control method and the input output (I/O) data acquired used to train the ANFIS. The performance of the proposed strategy should be tested on a feasible case study.

V. CONCLUSION AND DIRECTIONS OF FUTURE RESEARCH

This study reviewed the advances made towards the utilization of microgrid systems with PV-Wind generation units with emphasis on the applicable microgrid control systems, strategies and principles. Based on the review conducted, it was noted that every control strategy has its merits and demerits as well as levels of efficiency and cost particularly if extended to different microgrid scenario. As a result of this, a Multilevel Microgrid Control System (MMCS) incorporating suitable artificial intelligence algorithms has been suggested as a viable solution to the control challenges of PV-Wind hybrid

generation microgrid system. The directions taken by research in this field include using artificial intelligence and other nature-inspired algorithms to optimize performance of a hybrid microgrid system. Further, the decentralized hierarchical control is increasingly becoming common.

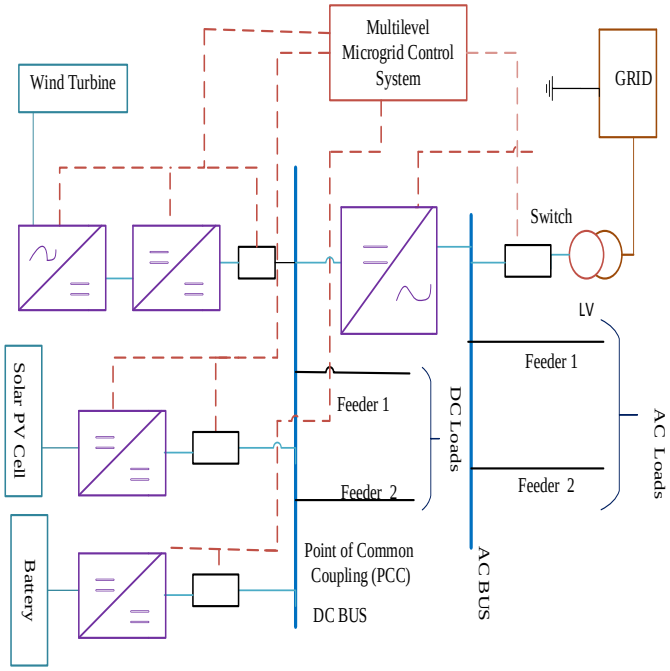


Fig.9 Proposed Microgrid Control System Layout

REFERENCES

[1] M. A. Aminu and K. Solomon, "A Review of Control Strategies for Microgrids," Advances in Research, SCIENCE DOMAIN international, Article no.AIR.25722-ISSN: 2348-0394, 2016.

[2] A. Bidram, V. Nasirian, A. Davoudi and F. L. Lewis, "Cooperative Synchronization in Distributed Microgrid Control", Advances in Industrial Control, Springer International Publishing AG, XVI, pp. 242, 2017.

[3] M. N. Ahmed, M. Hojabri, A.M. Humada, H. B. Daniyal and H.F. Frayyeh, "An Overview on Microgrid Control Strategies," International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-4 Issue-5, June 2015.

[4] D. E. Olivares, C. A. Cañizares, and M. Kazerani, "Trends in Microgrid Control," IEEE Transactions on Smart Grid, Vol. 5, No. 4, 2014.

[5] Y. Sawle1, S.C. Gupta1 and A. K. Bohre, "PV-wind hybrid system: A review with case study," Cogent Engineering, 2016.

[6] Y. Shen, "Microgrid Control Strategy Study and Controller Design Based On Model Predictive Control," Master of Science Thesis, University of Denver, March 2017.

[7] R. A. Badwawi1, M. Abusara1 and T. Mallick, "A Review of Hybrid Solar PV and Wind Energy System," Smart Science, Vol. 3, No. 3, pp. 127-138, April 2015.

[8] O. Palizban and K. Kauhaniemi, "Microgrid Control Principles in Island Mode Operation," carried out in Smart Grids and Energy Markets (SGEM), IEEE Conference Paper, 2013.

[9] P. Borazjani, N.I.A. Wahab, H. B. Hizam and A. Bt Che Soh, "A Review on Microgrid Control Techniques," IEEE Innovative Smart Grid Technologies - Asia (ISGT ASIA), 2014.

[10] K. A. Jada, H. M. Karkar and I. N. Trivedi, "A Review of Microgrid Architectures and Control Strategy," Springer, The Institution of Engineers (India), 2017.

[11] Z. Chen, K. Wang, Z. Li and T. Zheng, "A Review on Control Strategies of AC/DC Micro Grid," IEEE, 2018.

[12] D. Ton, "Microgrid Controls and Management Systems," Power Systems Engineering Research and Development, Office of Electricity Delivery & Energy Reliability, USA, July 2015.

[13] M. A. Hossain, H. R. Pota, W. Issa and M. J. Hossain, "Overview of AC Microgrid Controls with Inverter-Interfaced Generations," Energies Article, 10, 1300, 2017. Available at www.mdpi.com/journal/energies. [Accessed: 28/3/2018].

[14] L. Meng, Q. Shafiee, G. F. Trecate,H. Karimi, D. Fulwani, X. Lu and J.M. Guerrero, "Review on Control of DC Microgrids and Multiple Microgrid Clusters," IEEE Journal of Emerging and Selected Topics In Power Electronics, Vol. 5, No. 3, September 2017.

[15] T.M. Reshma and J. Deepu, "Droop Control Method for Parallel DC Converters used in Standalone PV-Wind Power Generating System," International Journal of Research in Engineering and Technology, Vol. 4 Issue 10,pp. 297-302, October 2015.

[16] Y. LI and F. NEJABATKHAH, "Overview of control, integration and energy management of microgrids," Springer, J. Mod. Power Syst. Clean Energy, 2(3):212–222, August 2014.

[17] L. I. Minchala-Avila, L. E. Garza-Casta˜n´on, A. V. Martinez, and Y. Zhang, "A review of optimal control techniques applied to the energy management and control of microgrids," Elsevier-ScienceDirect, Procedia Computer Science issue 52, pp. 780 – 787, 2015.

[18] U.S. Department of Energy (DOE), "Enabling Modernization of the Electric Power System," Quadrennial Technology Review: An Assessment of Energy Technologies and Research Opportunities, Chapter 3, September 2015.

[19] T. Morstyn, B. Hredzak, R.P. Aguilera and V. G. Agelidis, "Model Predictive Control for Distributed Microgrid Battery Energy Storage Systems," IEEE, February 2017.

[20] Y. DU, W. PEI, N. CHEN, X. GE and H. XIAO, "Real-time microgrid economic dispatch based on model predictive control strategy," Springer, J. Mod. Power Syst. Clean Energy, 5(5), pp. 787–796, February 2017.

[21] T. Wang, X. He and T. Deng, "Neural networks for power management optimal strategy in hybrid microgrid, The Natural Computing Applications Forum 2017, Springer, 2017.

[22] N. Borges, J. Soares and Z. Vale, "A Robust Optimization for Day-ahead Microgrid Dispatch Considering Uncertainties," Elsevier-ScienceDirect, IFAC PapersOnline 50-1, 3350-3355, 2017.

[23] M. Jayachandran and G. Ravi, "Design and Optimization of Hybrid Micro-Grid System," Elsevier-ScienceDirect, Energy Procedia, 117, pp. 95–103, 2017.

[24] V. Lavanya and N. S. Kumar, "A Review: Control Strategies for Power Quality Improvement in Microgrid," INTERNATIONAL JOURNAL of RENEWABLE ENERGY RESEARCH, Vol.8, No.1, March, 2018.

[25] GE Digital Energy, "Grid IQ™ Microgrid Control System: Optimization Solution for Permanently Islanded or Grid-Connected Microgrids," 2012, Available at GEDigitalEnergy.com. [Accessed: 30/3/2018].

[26] M. Jansen, "Microgrid Plus-A networked, easy to apply power flow control system for stable integration of intermittent renewable generation into microgrids," ABB Australia Pty. Ltd, Available at www.abb.com [Accessed: 1/4/2018].

[27] F. Katiraei, A. Zamani, and R. Masiello, "Microgrid Control Systems: A Practical Framework," IEEE Power & Energy Magazine, pp.110-116, 2017.

[28] R. Jain and A. Arya, "A Comprehensive Review on Micro Grid Operation, Challenges and Control Strategies," ACM. ISBN 978-1-4503-3609-3/15/07, 2015.

[29] W. Huang, M. Lu and L. Zhang, "Survey on Microgrid Control Strategies," Elsevier-ScienceDirect, Energy Procedia, 12, pp. 27–30, September 2011.

[30] D. Akinyele, J. Belikov and Y. Levron, "Challenges of Microgrids in Remote Communities: A STEEP Model Application," Energies-MDPI Article, 11, 432, February 2018.

[31] I Szeidert, I. Filip and O. Prostean, "Issues regarding the modelling and simulation of hybrid micro grid systems," IOP Conf. Series: Materials Science and Engineering 106 012028, 2016.

[32] K. Patel and S. Singh, "An Overview of DC-Microgrid Architecture with Renewable Energy Resources," International Journal of Advance Engineering and Research Development (IJAERD), 2015.

Review of Discrete Element Modelling in Optimisation of Energy Consumption of a Single-Toggle Jaw Crusher

Peter Ndung'u Mwangi, Onesmus Mutuku Muvengei and Thomas Ochuku Mbuya

Abstract—This review paper has focused on optimisation of energy consumption during comminution process within the crushing chambers of a single toggle jaw crusher. Comminution has always been associated with operational efficiencies of below 10% due to energy dissipation during size-reduction through noise, heat and vibrations. Optimisation techniques such as altering the feed material characteristics through heat treatment have been proposed. However, such techniques increase the overall cost of aggregate production. Therefore, the most appropriate technique is optimising the jaw crusher dimensions such as the reduction ratio, throw and toggle speed using Genetic Algorithms (GA) and Discrete Element Modelling (DEM). The Rose and English model has been proposed for optimisation and is used together with GA. Even though the GA will reveal the optimal design parameters, it does not account for stresses, deflections and physical properties of the material and jaw crusher. This gives a strong incentive to use Discrete Element Method (DEM) in performance optimisation of the jaw crusher. DEM modelling has the capability of predicting the machine's power draw and process parameters. DEM models that have been proposed included Bonded Particle Model (BPM), Particle Replacement Model (PRM) and Fast Breakage Model (FBM). Of the three, BPM was found to be superior in compression induced fractures. PRM is suitable for impact crushing while FBM is ideal for flow-induced breakage.

Keywords—Bonded Particle Modelling, BPM, Comminution, Discrete Element Modelling, DEM, Efficiency, Jaw crusher optimisation, power consumption, size reduction,

I. INTRODUCTION

IN quarries, jaw crushers are at the first stage of comminution processing for the manufacture of aggregates. There are numerous types of crushers which have been developed over the years for various materials and required products. Experimental approaches have always been used to optimise and customize crushers. However, optimisation using experimental techniques, such as monitoring of mass flow, where particles break and how they break can be difficult. Basic input and output information such as feed size, feed rate, desired product size and power draw can be measured [1].

Comminution is the process of breaking rock particles into smaller fragments (aggregates). Aggregates are used as building materials in modern society. Naturally, the comminution process has been associated with low efficiency, which is typically less than 10%. This dissipative nature arises from random application of forces inside the machine and between the particles. Therefore, for the efficiency of the comminution process to be improved, the process has to be understood [2].

This gives a strong incentive to improve the production of high quality aggregates at low cost and minimal environmental footprint. To design and create more optimized jaw crushers, better evaluation tools need to be implemented in the design process. Experimental techniques require development of physical prototypes which are expensive and the test process is cumbersome. This has forced researchers to turn to virtual prototypes during design and evaluation [3].

Discrete Element Method (DEM) is one of the most robust computational technique that allows particle flows in crushing equipment to be simulated. The major advantage of DEM modelling is that it has the ability to simulate how particles flow through the equipment under controlled conditions and provide detailed predictions of specific outputs [4]. DEM provides a detailed fracture process during comminution. Simple geometries and single particles have been studied in the past by various authors. Usually, rock particles are formed by creating clusters of DEM particles which are bonded together.

II. DISCRETE ELEMENT METHOD

A. Background

DEM simulation was first proposed by Cundall and Strack [5] as a numerical model capable of describing the mechanical behaviour of assemblies of discs and spheres. It is based on the use of an explicit numerical scheme where interaction of particles is monitored contact-by-contact and motion of the particles modelled particle-by-particle. An improvement was later made to DEM by introduction of Newton's second law of motion, linear hysteresis and use of contact models such as Hertz-Mindlin [3] [6].

DEM simulates discrete matter in a series of events known as time-steps. Once the particles are generated, the interaction between them is controlled using contact models and forces acting on all particles is calculated. Newton's second law of motion is then applied and the position of all particles can be calculated for next time-step [3]. DEM calculation loop is as shown in Fig 1.

B. Hertz-Mindlin Contact Model

In EDEM software, the no-slip Hertz-Mindlin contact model has two components; the normal force component and tangential force component. During simulation, the particles are

Peter Ndung'u Mwangi, Department of Mechanical Engineering, JKUAT (phone: +254702688000; e-mail:peter.mwangi@jkuat.ac.ke, mwangipeter423@gmail.com).

Onesmus Mutuku Muvengei, Department of Mechanical Engineering, JKUAT (e-mail: mmuvengei@eng.jkuat.ac.ke).

Thomas Ochuku Mbuya, Department of Mechanical Engineering, University of Nairobi (e-mail: tmbuya@uonbi.ac.ke).

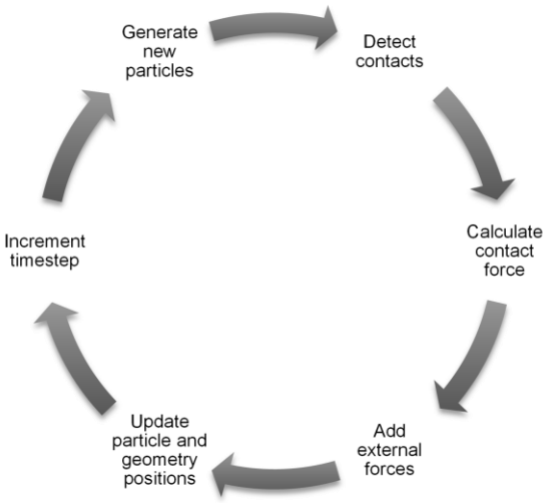


Fig. 1: Calculation loop used in DEM

allowed to overlap and the amount of overlap δx as well as the normal and tangential velocities at each contact point determine the collisional forces. A linear spring-dashpot model as shown in Fig 2 is used for the breakage algorithm. The

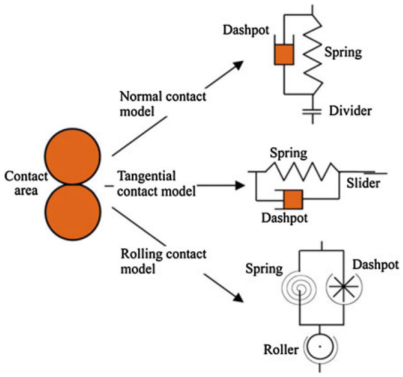


Fig. 2: Contact model between two spheres [7]

normal force is given by:

$$F_n = -k_n \Delta x + C_n v_n \mathbf{N} \tag{1}$$

Where,

$$C_n = 2\lambda\sqrt{m_{ij} k_n} \tag{2}$$

$$\lambda = \frac{-ln(\epsilon)}{\sqrt{\pi^2 + ln^2(\epsilon)}} \tag{3}$$

$$m_{ij} = \frac{m_i m_j}{m_i + m_j} \tag{4}$$

and

C_n= Damping coefficient of normal force in Ns/m.
k_n= Normal stiffness in N/m
ϵ= Coefficient of restitution.

Equation (1) has a linear spring to provide the repulsive force and a dashpot to dissipate a proportion of the relative motion. The term ϵ represents the coefficient of restitution which is defined as the ratio of the post-collisional to pre-collisional normal component of the relative velocity. In equation 4, the term m_{ij} represents the reduced mass of particles i and j with masses m_i and m_j respectively.
The tangetial force is given by:

$$F_t = min\left\{\mu F_n \sum k_t v_t \Delta t + C_t v_t\right\}\mathbf{N} \tag{5}$$

F_t and v_t are defined in the tangent plane at the contact point. The summation term in equation (5) represents an incremental spring that stores energy from the relative tangential motion and corresponds to the elastic tangential deformation of the contacting surfaces. The dashpot produces dissipation of energy in the tangential direction and represents plastic deformation of the contact in the tangential direction.

C. Calculation of stresses in DEM

DEM is only applicable for modelling the mechanical behaviour of discrete material such as rock and sand. In continuum mechanics, the stress-distribution cannot be obtained directly within the granular assembly. Homogenization or micro-macro averaging technique is used to calculate the stress tensor [7]. For an assembly of granular materials within a measurement volume (V), the stress tensor is defined as[8]:

$$\overline{\sigma_{ij}} = -\left(\frac{1-n}{\sum_{N_P} V^p}\right)\sum_{N_p}\sum_{N_p} |x_i^c - x_i^p| n^{(c,p)} F_j^c \tag{6}$$

In equation (6), the summation is taken over N_p particles whose centroids are located within the measurement volume V, n refers to the porosity of the measurement volume, V^p is the volume of a single particle; x_i^p is the location of the particle centroid; N_c is the number of contacts surrounding a single particle and $n^{(c,p)}$ is the unit normal directed from the particle center to the contact location x_i^c. The force acting at the contact as shown in Fig 3 is F_j^c.

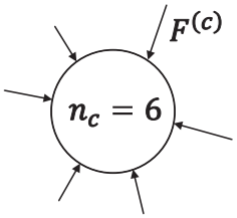


Fig. 3: Forces acting on a particle

1) Particle Stress Indexes: Grain crushing is induced by deviatoric stresses and the octahedral shear stress can be used to quantify a characteristic stress [9]. The stress index 1 based on the octahedral shear stress is as shown in equation (7).

$$\sigma_{i(1)} = \frac{1}{3}\left[(\sigma_1 - \sigma_3)^2 + (\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2\right]^{0.5} \tag{7}$$

Where, σ_1 is the major principal stress and σ_3 is the minor principal stress. In 2D, the intermediate stress is 0. Particles with just a few contact points can break because forces acting on them are not well distributed [10]. In addition, the crushing

of an individual particle is controlled by the largest contact force [11].

Stress index 2 is given by equation 8:

$$\sigma_{i(2)} = \sigma_t = \frac{2F_{max}}{\pi LD} \tag{8}$$

where F_{max} is the maximum normal contact force acting on the particle, D is the particle diameter and L is the thickness of the disk. Stress Index 3 can be defined as;

$$\sigma_{i(3)} = \frac{\sigma_1 - 3\sigma_3}{2} \tag{9}$$

where σ_o is the strength of the particle whose diameter D_o is known. The strength σ_o can be obtained by measuring the maximum contact force that can be carried by a single particle of a given diameter [12] [13].

2) Particle Strength: For a particle of diameter D, its strength can be defined as;

$$\sigma_{s(D)} = \sigma_o \left(\frac{D}{D_o}\right)^{\frac{-3}{m}} \tag{10}$$

D. Particle Breakage Models Used in DEM

Zheng and Tannant [14] modelled the grain breakage criteria of sand crushing using DEM. Laboratory tests were conducted so as to calibrate the discrete element model. The breakage grain criteria was observed using PFC2D software. Three methods of simulating grain breakage were proposed; Bonded Particle Model, Fast Breakage Model and Particle Replacement Model (or Population Balanced Model).

1) Bonded Particle Model (BPM): The bonded particle model (BPM) directly mimics the behaviour of cemented granular materials. A BPM particle is made up of coarse and fine particles which end up forming an irregular or regular shape [15]. The fine particles act as the cement (connections) which break when applied load goes beyond the pre-defined critical level [6]. BPM has been used to carry out investigations on the fracture behaviour of single rock particles. Potyondy demonstrated that this model had the ability of replicating the strain softening of granite in both confined and unconfined conditions [8].

The BPM model was first published by Potyondy and Cundall [8]. The particles bonded together are known as fraction particles while the resulting cluster is known as a meta-particle. A bi-modal particle size distribution is proposed by Johannes [3] as it forms realistic bonds. A high packing density should be achieved so as to eliminate the problem of mass-conservation due to the fact that the clustered rock particle might not be able to attain full solid density. During fragmentation, the bulk density of the meta-particle will change as new particles are formed. Unlike Particle Replacement Model, BPM is dependent on particle flow dynamics with the crushing chamber. Fig 4 depicts the forces and moments that act on a single BPM bond. These forces are given by:

$$\begin{aligned} \Delta F_b^n &= -k_b^n A_b v^n \Delta t \\ \Delta F_b^s &= -k_b^s A_b v^t \Delta t \\ \Delta M_b^n &= -k_b^s J \omega^n \Delta t \\ \Delta M_b^s &= -k_b^n J \omega^t \Delta t \end{aligned} \tag{11}$$

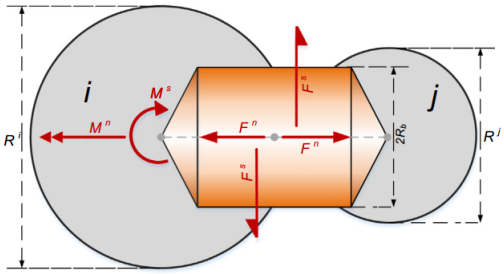


Fig. 4: Illustration of BPM forces and moments

where,

$$\begin{aligned} A_b &= \pi R_b^2 \\ J &= \frac{1}{2}\pi R_b^4 \\ I &= \frac{1}{4}\pi R_b^4 \end{aligned} \tag{12}$$

The terms k_b^s, k_b^n are shear and normal bond stiffness respectively, $v^n, v^t, \omega^n, \omega^t$ are normal and tangential velocities, and the normal and tangential angular velocities respectively. I, J and A_b are moment of inertia, polar moment of inertia and the cross-sectional area of the bond respectively. R_b is the bond radius and is a dependent of the smallest sphere in contact. Moment M_b and force F_b act in the normal direction (M_b^n and F_b^n) and shear direction(F_b^s and M_b^s) as shown in 11.

For fracture to occur, the applied normal and shear stresses must be greater than the critical stresses, σ_c and τ_c which are user-defined. The most crucial part of BPM is sphere size selection which make up the meta-particle. They could be normal distributions [16], bimodal distributions [3] or mono-dispersed [17].

Quist [3] used the BPM in modelling the crushing process within the cone-crusher. The simulation findings were compared against experimental findings. It was asserted that the particle flow behaviour through the crushing chamber looked realistic in comparison to what was observed from experiments. The simulations also revealed that the BPM approach captures the breakage process and the particles were able to react to different compressive breakage modes. The results obtained revealed that BPM simulation can be used to predict the mass flow and pressure but more work has to be performed on the post-processing for accurate power draw and size distribution predictions. The Single Particle Breakage (SPB) test was used as it gave more insight on how compression mode influences rock fracture. Legendre [2] used the SPB test and BPM model to assess the energy efficiency of a jaw crusher. The simulation was found to follow the operation of the real machinery during particle breakage. Potyondy [8] adopted the BPM in modelling of granite rock. It was concluded that BPM allows the cracks to form, interact with each other and coalesce into macroscopic fractures. The model was found to reproduce many features of rock behaviour such as elasticity, fracturing and strength increase with confinement. Jimenez [6] observed that BPM is capable of reproducing several aspects of particle breakage such as the force-deformation profile and how the particles

interact with the crushing walls. Diego [18] used BPM to simulate the mechanical behaviour of jointed rock masses. BPM was used to represent the intact rock material. It was concluded that this technique is suitable for simulating rock properties (fracture toughness, fracture strength, dilation angle, peak strength, etc.) which influence the comminution process.

2) Fast-Breakage Model: This technique has been used by Potapov and Campbell [19], Paluszny et al. [20] and in simulation of comminution equipment [21]. This method uses polyhedral elements to form DEM particles. It is a type of particle-replacement model. FBM is an instantaneous breakage model that uses the Laguerre-Voronoi tessellation to segment 3D polyhedral when the total energy after first collision is greater than the fracture energy. Bonds between individual particles are described using the linear hysteresis contact model as shown by Wu et al. [22]. The probability of breakage *P(E)* is described by the expression originally proposed by Vogel and Peukert [23]:

$$P(E) = 1 - exp[-S\left(\frac{d_i}{d_{i,ref}}\right)E_{cum}] \tag{13}$$

and,

$$\begin{aligned} E_{cum} &= E'_{cum} + E - E_{min} \\ E_{min} &= e_{min,ref}\left(\frac{d_{i,ref}}{d_i}\right) \end{aligned} \tag{14}$$

where, E'_{cum} is the accumulated energy in the particle just before the stressing event and E is the total specific energy of the collision during an impact event. S, $d_{i,ref}$ and $e_{min,ref}$ are model parameters which describe the amenability of the material to breakage, the reference size and the minimum energy for breakage respectively. The particle size is represented by d_i.

Jimenez [6] concluded that FBM describes the interaction between a particle and particle bed during impact crushing. In addition, this technique was found to have the ability to conserve mass and generate irregularly shaped fragments, hence making it suitable for simulations of large-scale comminution systems. However, FBM had limitations in describing the measured force-deformation profile from SPB as well as the progeny size distribution. Potapov [19] used FBM to study rock fracture and asserted that this method can be used to predict the size distribution for fragments larger than three element sizes. In addition, FBM is suitable for simulations involving flow-induced breakage of granular systems. Brosh et al. [24] applied FBM in CFD-DEM simulation of a size-distribution system. It was concluded that FBM is suitable for predicting size-reduction in systems where size-reduction is significant. However, FBM was found to require long computational time for simulations with large number of particles.

3) Particle Replacement Model: Particle Replacement Model (PRM), also known as Population Balanced Model (PBM), was first proposed by Cleary [25]. This model is based on the principle of replacing parent particles with progeny particles once they are subjected to a load greater than the critical strength [3] [6]. The set of progeny particles are of a predetermined size and shape. This model is analysed by Barrios et al. [26] using DEM software. Generation of progeny particles stops when the breakage process contains particles smaller than the minimum size. PRM does not take into account the particle flow dynamics as new particles are introduced at the same position as the broken parent particle i.e. particle dynamics are decoupled from the parent particle. Huiqi Li [27] explored the comminution process within the cone crusher using PRM. The results obtained from the simulations were compared against experimental findings. The effect of varying the Closed Side Set (CSS) and eccentric speed was investigated. Simulation results concurred with experimental data. However, the effect of fracture strength on energy consumption was not investigated. Jimenez [6] affirmed that PRM provided an ideal visualisation of size distribution of progeny particles from both SPB and particle bed experiments. However, it was not able to describe the measured force-deformation profile and there was no stochastic description of breakage probability. PRM gave poor descriptions of how the rock model interacted with the particle bed.

4) Weaknesses and Strengths of DEM Particle Breakage Models: BPM model has been used to simulate comminution process and model rocks by various authors [2] [3] [6] [8] [18] [28] [29]. BPM has been observed to reproduce superior power draw and force-deformation predictions than PRM and FBM techniques. In addition, BPM has been able to vividly describe the interaction between modelled rocks and compression walls. The cracks generated by BPM are able to interact with each other hence producing a more realistic approach in size-reduction simulations in comparison to PRM. However, BPM has been found to be less suitable for systems where large particle populations are required as it demands high computational power[6]. On its own, BPM is not capable of giving firm conclusions regarding power draw and size distribution and hence post-processing and analysis has to be done using MATLAB [3].

FBM is suitable for simulating size-reductions that are flow-induced [19][24]. In addition, this technique describes the interaction between particles and machinery walls during impact crushing. FBM has the ability to generate irregularly shaped fragments and conserve mass during comminution. However, just like BPM, this technique requires massive computational power for simulations with large number of particles. The force-deformation profile was also not presented using FBM and progeny size distribution was found to be inferior to those generated using BPM [6].

PRM is another DEM technique that has been used in size-reduction simulations [6] [25], [26] [27]. PRM has is suitable for simulations where impact crushing is dominant and large number of particles are to be simulated. PRM does not provide a good description of breakage probability and force-deformation profile.

Cabiscol et al. [30] used the multi-sphere feature within EDEM software to model uniaxial-compressed cylindrical tablets. It was discovered that the shape and edges of DEM tablet models affected the packing fraction hence needed to be done at high precision. Sliding and rolling friction were identified as the most sensitive parameters through a pouring experiment. Calibration for discrete element simulation was carried out using the data obtained from tumbling drum and

pouring tests i.e. the coefficient of restitution and Young's modulus. However, DEM was found to have the following limitations:

1) It was not capable of extrapolating input parameters for another system.
2) Some of the input parameters lacked physical meaning.

Metta et al. [31] used PBM to simulate a comill process. Experiments where the impeller speed was varied were conducted. It was observed that the impeller speed affected the particle size distribution and process parameters such as time required to attain a steady state and the hold up time. Therefore, an iterative algorithm that used mechanistic information from DEM and process variables from experiments was proposed. A breakage kernel was also created from the simulation.

III. POWER CONSUMPTION IN SINGLE TOGGLE JAW CRUSHERS

The single toggle jaw crusher is a comminution machine which has been associated with low energy efficiencies [2][32]. Legendre and Ron [2] demonstrated that the energy efficiency of a single toggle jaw crusher can be improved by either the pre-treatment of the feed material or altering the design parameters. Pre-treatment of feed material, however, was disregarded as it might increase the overall cost of comminution. Therefore, finding the optimal design machine parameters would be more realistic. Use of Evolutionary Algorithms and DEM was recommended as this method accounts for particle dynamics within the crushing chamber and will attempt to mimic an on-site jaw crusher [2]. Fig 5 shows the critical jaw crusher parameters that affect power consumption. Even

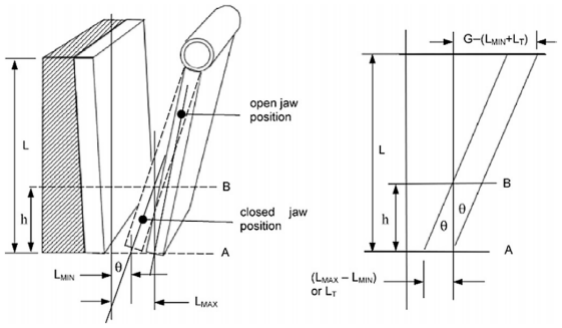


Fig. 5: Machine Parameters of a Single-toggle jaw crusher [2]

though jaw crusher design parameters such as the gape, throw, reduction ratio and operating speed affect power consumption, most authors have demonstrated that other factors such as weight of the swing jaw [32] [33] [34] and shape of the particle play an important role in power consumption [1] [35].

Bharule [33], Ramkrushna [34] and Sundar [36] investigated the effect of adding stiffeners to the swing jaw. Stiffeners increase the strength to weight ratio of the swing jaw. It was concluded that increasing the number of stiffeners will reduce the power consumption. Bharule [33], using CatiaV5R15, showed that stiffeners reduced energy consumption during comminution by 25%. Miller [37] designed a new crushing technique in which both crushing plates moved during crushing. This prototype reduced the fracture energy of the feed material but increased the cost of production i.e. the additional moving jaw required additional of extra components such as springs, bolts and machining/casting. The author reported that reduction ratio, material properties and production flow rate played a major role in power consumption.

A. Laws of Comminution Energy

In design, operation and control of comminution processes, it is necessary to correctly evaluate the comminution energy of solid materials [38]. Evaluating the comminution energy will enable engineers and manufacturers design efficient jaw-crushers and also predict the power draw of the machine. It has been observed that as the particle size decreases, the surface area of the particles increases. Therefore, the surface area before and after size reduction would indicate the amount of energy consumed during comminution[39]. Therefore,

$$dE = k[S^n dS] \tag{15}$$

where, E is the energy used during comminution, S is the surface area, k is a constant and a function of the crushing strength of the rock, and n is an exponent which has been determined by Rittinger, Kick and Bond [39].

Size reduction energy is expressed as a function of the feed size and product size. Four laws have been developed to model comminution energy; Rittinger's Law, Kick's Law, Bond's Law and Holmes' Law.

a) Rittinger's Law: Rittinger assumes that the energy consumed is directly proportional to the new surfaces created. Therefore, specific surface area is inversely proportional to the particle size. The specific comminution energy E/m is given by:

$$\frac{E}{m} = C_R(S_P - S_F) \ \textbf{[kJ/kg]} \tag{16}$$

Where, m is the particle mass, S_p and S_f are specific surface areas of product and feed respectively, and C_R is a constant that depends on the characteristics of the material.

b) Kick's Law: Kick law states that the energy required for comminution is related only to the ratio between feed size and product size.

$$\frac{E}{m} = C_K \ln\left(\frac{x_f}{x_p}\right) \ \textbf{[kJ/kg]} \tag{17}$$

Where x_p and x_f are the particle sizes of product and feed respectively while C_k is a constant. Equation (17) can be derived by assuming that the strength is independent on particle size and the energy for size reduction is proportional to the volume and size reduction is constant at each stage of comminution. This is in contradiction to Griffiths weakest link theory which mentions that as particles get smaller, they become more resistant to fracture [33].

c) Bond's Law: Bond approach has always been used in predicting the energy consumption during the size-reduction process. Bond Work Index coefficients cover almost the entire range of particles that are to be processed using comminution machinery[40]. Bond suggests that any comminution process

can be considered to be an intermediate stage in the breakdown of a particle of infinite size to an infinite number of particles of zero size. Bond's theory states that the total work useful in breakage is inversely proportional to the square root of the size of the product and feed particles:

$$W = W_i\left(\frac{10}{\sqrt{P_{80}}} - \frac{10}{\sqrt{F_{80}}}\right) \text{ [kWt/hr]} \quad (18)$$

Where, W is the work input, W_i is the Bond's work index, P_{80} is the Product size in microns at which 80% passes through the sieve[μm] and F_{80} is the Feed size in microns at which 80% passes through the sieve[μm]. From equation 15, the values of n depend on the type of application and are as follows:

- n= -2 (Rittinger)
- n= -1 (Kick)
- n= -1.5 (Bond)

Rittinger's expression n=-2 is applicable for coarse size reduction while for Kick n=-1 is suitable for finer size reductions. Bond's value of n=-1.5 covers almost the entire range of particles and has been accepted universerly [39].

However, the laws of comminution have a few weaknesses; they are not suitable for evaluation of primary crushing equipment processing large particle sizes, they only fit experimental data over a limited range of variables and suitable for only certain cases and Bond's work results in a conservative overdesign of the crushing plant. To counter these weaknesses, Single Particle Breakage analysis is currently being adopted to relate fracture energy and product size distribution to a materials property. There is enough evidence that a link exists between rock properties, energy consumption and performance of jaw crusher [32].

B. Single Particle Breakage

The mechanism of size reduction of solids is based on the fracture of a single particle and its accumulation during comminution operations[38]. The elastic strain energy, E $(Joules)$, input to a sphere up to the instant of fracture is given by:

$$E = 0.832\left(\frac{1-\nu^2}{Y}\right)^{\frac{2}{3}}(D)^{-\frac{1}{3}}(P)^{\frac{5}{3}} \quad (19)$$

The particle is subjected to compressive point loading. The tensile strength S (Pa) of the specimen is given by:

$$S = \frac{2.8P}{\pi D^2} \quad (20)$$

Substituting equation (20) to equation (19), the specific fracture energy E/m[**J/kg**] is given by:

$$\frac{E}{m} = 0.897\rho^{-1}(\pi)^{\frac{2}{3}}\left(\frac{1-\nu^2}{Y}\right)^{\frac{2}{3}}(S)^{\frac{5}{3}} \quad (21)$$

Where ρ is the density of the sphere in kg/m^3. When two spherical particles, 1 and 2, collide with each other, the maximum stress S_{max} generated inside the particles is expressed by a function of particle size, x, relative velocity ν (m/s) and mechanical properties:

$$S_{max} = 0.628\left(\frac{m_1 m_2}{m_1+m_2}\right)^{0.2}\mu^{0.4}\left(\frac{2}{D_1}+\frac{2}{D_2}\right)^{0.6}\left(\frac{1-\nu_1^2}{Y_1}+\frac{1-\nu_2^2}{Y_2}\right) \quad (22)$$

Where m_1 and m_2 are masses of particles 1 and 2 respectively in kg, μ is the relative velocity between the two particles in m/s, D_1 and D_2 are diameters of particle 1 and particle 2 respectively in m, ν_1 and ν_2 are poisson's ratio of particle 1 and 2, and Y is the Young's modulus ni MPa.

There is a direct relationship between fracture toughness and specific comminution energy [33]. From previous research [33] [37] [41], it is clear that a materials's reaction to applied load is dependent on size, shape and conditions under which the load is applied. Donovan [33] and Miller [37], demonstrated that bigger particles are easier to fracture than smaller ones. The shape of the feed material affects; the shear strength, realistic voidage(how the particles pack in the crushing chamber) inside the jaw crusher and, how and when the particles break [1][25].

C. Genetic Algorithms

Genetic algorithms (GAs) are numerical optimization algorithms which are inspired by both natural selection and natural genetics. A typical algorithm uses three operators, selection, crossover and mutation (chosen in part by analogy with the natural world) to direct the population (over a series of time steps or generations) towards convergence at the global optimum [42]. There are many types of Genetic Algorithms which come with different representations. The two most important GAs are the Simple Genetic Algorithm (SGA) and the Differential Evolution (DE) methods. It has been shown that a SGA code has a better potential of solving power consumption than a DE code [2]. Lee [43] used a genetic algorithm to optimize a jaw crusher performance theoretically. Theoretical efficiencies of 30%-40% were obtained. It should be noted that GA codes do note take into account the weight of the jaw plate, the friction between moving parts and vibrations induced during comminution. Therefore, GA codes should be used together with DEM for better optimisation.

The Rose and English [39] jaw crusher model should be adopted for flow rate and power consumption optimisation [2].

$$\dot{W} = 0.01195 W_i Q_A\left[\frac{\sqrt{G}-1.054\sqrt{L_{MIN}+L_T}}{\sqrt{G}\sqrt{L_{MIN}+L_T}}\right]\text{[kWh/t]} \quad (23)$$

Where, W_i is Bond's Work Index, Q_A is the actual flow rate in t/h, G, L_{MIN} and L_T are gape, closed side set and throw in m respectively. The maximum flow rate Q_A of a jaw crusher is calculated as:

$$Q_M = 60 L_t v_{cr} w(2L_{min}+L_t)\left(\frac{R}{R+1}\right)\rho f(P_k) f(\beta) S_c \text{ (t/h)} \quad \text{(A.1)}$$

The flow is affected by empirical functions $f(\beta)$ and $f(P_k)$. These functions are defined within a certain range; $0 \geq P_k \geq 1.2$ and $0 \geq \beta \geq 0.8$. The term S_c is a surface characteristic parameter related to the jaw crusher and has a range of $0.5 \leq S_c \leq 1$ [39].

The actual flow (Q_A) is the jaw crusher's real flow and is a function of operating velocity. It is given by:

$$Q_A = Q_M \frac{v}{v_{cr}} \text{ if } v \leq v_{cr} \text{ [t/h]} \quad (24)$$

and,

$$Q_A = Q_M \frac{v_{cr}}{v} \text{ if } v > v_{cr} \text{ [t/h]} \quad (25)$$

The critical velocity v_{cr} is calculated as[39];

$$v_{cr} = \sqrt{4414.5\left(\frac{G-(L_{min}+L_t)}{L}\right)\left(\frac{1}{L_t}\right)}\ [\textbf{rpm}] \quad (26)$$

The reduction ratio R is calculated as:

$$R = \frac{G}{Lmin} \quad (27)$$

D. Energy Efficiency Calculations

Legendre [2] approximated the energy consumed during comminution by measurement of surface area before and after size reduction of the feed.

$$\frac{dE}{dS} = kS^n \quad (28)$$

Where,

E= Indicator of the energy used in the comminution procedure.
S= The surface area before and after size reduction
K and n= Constants related to the crushing strength of the rock.

Tromans [44] defined the maximum ideal limiting efficiency based on compressive loading comminution machine, which generated a stress distribution inside a single particle as a result of a central crack flaw.

$$\eta_{limit} = \left[\frac{\sigma_x^2 - 2\nu(\sigma_x\sigma_y+\sigma_x\sigma_z)}{(\sigma_x^2+\sigma_y^2+\sigma_z^2)-2\nu(\sigma_x\sigma_y+\sigma_x\sigma_z+\sigma_y\sigma_z)}\right]\times 66\% \quad (29)$$

$$\sigma_x = \sigma_y = \frac{P}{\pi D^2}\left(\frac{6(1-2\nu)}{2+\sqrt{2}}\right) \quad (30)$$

$$\sigma_z = -\frac{P}{\pi D^2}\left(12-\frac{3}{\sqrt{2}}\right) \quad (31)$$

Where,

P= Loading Force, N
D= Particle Diametre, m
$\sigma_{x,y,z}$= Stresses along x,y and z axes
ν= Poissons ratio,-,typically ranging from 0 to 0.5

Tromans and Meech [44] showed that the increase in surface energy per unit of mass from particle fracture can be approximated by:

$$\Delta S_{En} = \frac{\gamma A_{Created}}{m} \approx \frac{6F_r\gamma}{\rho}\left(\frac{1}{D_{f_{Sm}}}-\frac{1}{D_{i_{Sm}}}\right)[\textbf{J/kg}] \quad (32)$$

Where,

γ = Fracture energy in J/m^2
$A_{Created}$ = New surface area created in m^2
$D_{i_{Sm}}$ and $D_{f_{Sm}}$= Initial and final surface mean diameter of particles in m respectively.
F_r = Dimensionless surface roughness factor that corrects for non-sphericity $[1 \leq F_r \leq 3]$

Assuming that the performance of the jaw crusher is ideal, there will be a continuous steady-state comminution process whose power consumption is given by:

$$\Delta S_{Power} = \Delta S_{En}(\textbf{kJ/kg}) \times 1(\textbf{kg/s})[\textbf{kW}] \quad (33)$$

Equation (33) assumes a normalised flow of 1 kg/s hence does not represent the actual power draw of a real comminution process. However, it is useful to monitor the influence of the machinery design parameters during comminution. Combining equation (32) and equation (33), the equipment total energy efficiency can be obtained by:

$$\eta_{EnergyTotal} = \frac{\frac{\Delta S_{En}}{E_{equipment}}}{\eta_{limit}} = \frac{\eta_{Classical}}{\eta_{limit}} \quad (34)$$

During comminution, the compressive strength, fracture toughness and tensile strength have the greatest influence on energy consumption. Korman observed that larger sized product particles were obtained when the uni-axial compressive strength of the material increased [45]. Kirankumar [46] explained the optimisation of a jaw crusher by modifying the frame, redesigning the crushing chamber and re-organizing the crushing parameters such as size reduction processes. However, high reduction ratios result in poorer shaped particles and higher production costs. Therefore, optimal reduction ratios should be used in jaw crusher designs.

IV. CONCLUSION

Discrete element modelling is a robust method in simulating the comminution process within the crushing chambers of comminution machines. From the literature, most researchers have used Bonded Particle Model and Population Balanced Model for simulation purposes. The BPM model is suitable for simulating size-reduction processes in machines such as the cone-crushers and jaw crushers where the dynamics of crushed particles are crucial in overall machine performance. On the other hand, PBM is suitable for milling simulations. Calibration of a DEM simulation requires high precision in particle shape, material property (such as the coefficient of restitution, Young‘s Modulus, rolling friction, etc.) and operational parameters such as the toggle speed, feed rate and flow rate. It has also been observed that authors such as Legendre, Khanal and Donovan did not take into account the eccentricity of the swing-jaw during comminution. Neglecting this parameter affected the end prediction of power consumption as it deviates from the realistic nature of the jaw crusher.

A single toggle jaw crusher has multiple variables that affect the overall power consumption. According to Rose and English jaw-crusher model, power is a function of toggle speed, reduction ratio and throw. To find the optimal design parameters, Genetic Algorithms have been utilised by Legendre and Lee. Even though the GAs obtain the optimal design parameters, these codes do not take into account the machine and product mechanical properties. Therefore, they must be used together with DEM for realistic predictions.

The power consumption of a jaw crusher is also dependent on the feed material‘s fracture toughness, elastic modulus and reduction ratio. For instance, high reduction ratios are often associated with high power consumption. The reduction ratio also determined the capacity and throughput of the machine [39].

Stiffeners have also been used to increase the strength to weight ratio of the swing jaw [34] [41] [47]. It has been

demonstrated that stiffeners reduce power consumption by approximately 25%. However, the optimal number of stiffeners has not been determined, and the deflections and stresses associated to stiffeners have not been accounted for. The following gaps were identified:

1) The eccentricity of the swing jaw was neglected while simulating the DEM breakage process.
2) Product and process parameters such as particle shape and flow rate have been neglected by researchers.
3) The effect of stiffeners on stresses and deflections on the swing jaw have not been fully studied. In addition, the optimal number of stiffeners has not been determined yet.
4) Inter-particle interaction has not been included in simulations by researchers.

REFERENCES

[1] P. W. Cleary, "Industrial particle flow modelling using discrete element method," *Engineering Computations*, vol. 26, no. 6, pp. 698–743, 2009.

[2] D. Legendre and R. Zevenhoven, "Assessing the energy ef fi ciency of a jaw crusher," pp. 1–12, 2014.

[3] J. Quist, "Cone Crusher Modelling and Simulation," no. 1652, 2012.

[4] P. W. Cleary and M. D. Sinnott, "Simulation of particle flows and breakage in crushers using DEM: Part 1 - Compression crushers," *Minerals Engineering*, vol. 74, pp. 178–197, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.mineng.2014.10.021

[5] P. A. Cundall and O. D. L. Strack, "A discrete numerical model for granular assemblies," *Géotechnique*, vol. 29, no. 1, pp. 47–65, 1979.

[6] N. Jiménez-Herrera, G. K. Barrios, and L. M. Tavares, "Comparison of breakage models in DEM in simulating impact on particle beds," *Advanced Powder Technology*, 2017.

[7] T. Zhao, *Coupled DEM-CFD analyses of landslide-induced debris flows*. Singapore: Springer, 2017.

[8] D. O. Potyondy and P. A. Cundall, "A bonded-particle model for rock," *International Journal of Rock Mechanics and Mining Sciences*, vol. 41, no. 8 SPEC.ISS., pp. 1329–1364, 2004.

[9] J. R. Mcdowell, "DEM of triaxial tests on crushable sand," pp. 551–562, 2014.

[10] S. Lobo-Guerrero and L. E. Vallejo, "Crushing a weak granular material: experimental numerical analyses," *Géotechnique*, vol. 55, no. 3, pp. 245–249, 2005.

[11] A. R. Russell, D. Muir Wood, and M. Kikumoto, "Crushing of particles in idealised granular assemblies," *Journal of the Mechanics and Physics of Solids*, vol. 57, no. 8, pp. 1293–1313, 2009. [Online]. Available: http://dx.doi.org/10.1016/j.jmps.2009.04.009

[12] H. M. Y. Nakata, A. F. Hyde, "A probabilistic approach to sand particle crushing in the triaxial test," no. 5, pp. 567–583, 1999.

[13] D. Vallet and J. C. Charmet, "Mechanical behaviour of brittle cement grains," *Journal of Materials Science*, vol. 30, no. 11, pp. 2962–2967, 1995.

[14] W. Zheng and D. D. Tannant, "Computers and Geotechnics Grain breakage criteria for discrete element models of sand crushing under one-dimensional compression," *Computers and Geotechnics*, no. October, pp. 0–1, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.compgeo.2017.10.004

[15] B. Wang, U. Martin, and S. Rapp, "Discrete element modeling of the single-particle crushing test for ballast stones," *Computers and Geotechnics*, vol. 88, pp. 61–73, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.compgeo.2017.03.007

[16] S. Antonyuk, M. Khanal, J. Tomas, S. Heinrich, and L. Mörl, "Impact breakage of spherical granules: Experimental study and DEM simulation," *Chemical Engineering and Processing: Process Intensification*, vol. 45, no. 10, pp. 838–856, 2006.

[17] A. Spettl, M. Dosta, S. Antonyuk, S. Heinrich, and V. Schmidt, "Statistical investigation of agglomerate breakage based on combined stochastic microstructure modeling and DEM simulations," *Advanced Powder Technology*, vol. 26, no. 3, pp. 1021–1030, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.apt.2015.04.011

[18] D. Mas Ivars, *Bonded particle model for jointed rock mass*, 2010, no. January. [Online]. Available: http://kth.diva-portal.org/smash/record.jsf?pid=diva2:300557

[19] C. S. C. Potapov, Alexander V., "A three dimensional simulation of brittle solid fracture," *International Journal of Modern Physics*, vol. Volume 7, no. No. 5.

[20] A. Paluszny, X. Tang, M. Nejati, and R. W. Zimmerman, "A direct fragmentation method with Weibull function distribution of sizes based on finite- and discrete element simulations," *International Journal of Solids and Structures*, vol. 80, pp. 38–51, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.ijsolstr.2015.10.019

[21] J. Lichter, K. Lim, A. Potapov, and D. Kaja, "New developments in cone crusher performance optimization," *Minerals Engineering*, vol. 22, no. 7-8, pp. 613–617, 2009. [Online]. Available: http://dx.doi.org/10.1016/j.mineng.2009.04.003

[22] C. L. Wu, O. Ayeni, A. S. Berrouk, and K. Nandakumar, "Parallel algorithms for CFD-DEM modeling of dense particulate flows," *Chemical Engineering Science*, vol. 118, pp. 221–244, 2014.

[23] L. Vogel and W. Peukert, "From single particle impact behaviour to modelling of impact mills," *Chemical Engineering Science*, vol. 60, no. 18, pp. 5164–5176, 2005.

[24] T. Brosh, H. Kalman, and A. Levy, "Accelerating CFD-DEM simulation of processes with wide particle size distributions," *Particuology*, vol. 12, no. 1, pp. 113–121, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.partic.2013.04.008

[25] P. W. Cleary, "Large scale industrial DEM modelling," *Engineering Computations*, vol. 21, no. 2/3/4, pp. 169–204, 2004. [Online]. Available: http://dx.doi.org/10.1108/02644400410519730

[26] G. Barrios, L. Tavares, and J. Pérez-Prim, "DEM Simu-

lation of Bed Particle Compression Using The Particle Replacement Model," *Proceedings of the 14th European Symposium on Comminution and Classification*, no. September, pp. 59–63, 2015.

[27] H. Li, G. McDowell, and I. Lowndes, "Discrete element modelling of a rock cone crusher," *Powder Technology*, vol. 263, pp. 151–158, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.powtec.2014.05.004

[28] M. C. Weng and H. H. Li, "Relationship between the deformation characteristics and microscopic properties of sandstone explored by the bonded-particle model," *International Journal of Rock Mechanics and Mining Sciences*, vol. 56, pp. 34–43, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.ijrmms.2012.07.003

[29] M. Obermayr, K. Dressler, C. Vrettos, and P. Eberhard, "A bonded-particle model for cemented sand," *Computers and Geotechnics*, vol. 49, pp. 299–313, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.compgeo.2012.09.001

[30] R. Cabiscol, J. H. Finke, and A. Kwade, "Calibration and interpretation of DEM parameters for simulations of cylindrical tablets with multi-sphere approach," *Powder Technology*, 2017. [Online]. Available: https://doi.org/10.1016/j.powtec.2017.12.041

[31] N. Metta, M. Ierapetritou, and R. Ramachandran, "A multiscale DEM-PBM approach for a continuous comilling process using a mechanistically developed breakage kernel," *Chemical Engineering Science*, 2017. [Online]. Available: https://doi.org/10.1016/j.ces.2017.12.016

[32] B. A. Suresh, " Computer Aided Design and Analysis of Swing Jaw Plate of Jaw Crusher Computer Aided Design and Analysis of Swing Jaw Plate of Jaw Crusher Department of Mechanical Engineering National Institute of Technology," 2009.

[33] J. G. DONOVAN, *Fracture Toughness Based Models for The Prediction of Power Consumption, Product Size, And Capacity of Jaw Crushers*, 2003.

[34] R. S. More and G. H. Raisoni, "A Design & Analysis of Swing Jaw Plates of Jaw Crusher," vol. 3, no. 4, pp. 400–408, 2014.

[35] A. J. Magerowski, "Controlling Crushing Costs and Particle Shape."

[36] S. Sundar.V, "Optimum design and analysis of single toggle jaw crusher," vol. 8354, no. 3, pp. 194–203, 2014.

[37] R. B. Miller, "Designing A New Crushing Technique to Combine Impact and Compression Fracturing in A Rock Crushing Chamber," pp. 1–14.

[38] Y. K. Ã and N. Kotake, "Comminution Energy and Evaluation in Fine Grinding," vol. 12, no. 07, 2007.

[39] A. Gupta and D. Yan, *Introduction to Mineral Processing Design and Operation*, 2006.

[40] K. Schroe, *Production, Handling and Characterization of Particulate Materials*, 2016, vol. 25. [Online]. Available: http://link.springer.com/10.1007/978-3-319-20949-4

[41] G. Doktoringenieur, M. Khanal, M. Gutachter, T. Prof, A. B. Prof, and S. Luding, "Simulation of Crushing Dynamics of an Aggregate-Matrix Composite by Compression and Impact Stressings," no. January, 2005.

[42] D. Coley, "An introduction to genetic algorithms for scientists and engineers," p. 244, 1999.

[43] E. LEE, *Optimization of Compressive Crushing*, 2012.

[44] D. Tromans, "Mineral comminution: Energy efficiency considerations," *Minerals Engineering*, vol. 21, no. 8, pp. 613–620, 2008.

[45] T. Korman, G. Bedekovic, T. Kujundzic, and D. Kuhinek, "Impact of physical and mechanical properties of rocks on energy consumption of Jaw Crusher," *Physicochemical Problems of Mineral Processing*, vol. 51, no. 2, 2015.

[46] G. Kirankumar, "Optimization of Jaw Crusher," pp. 238–242, 2014.

[47] M. Khanal and C. T. Jayasundara, "Role of particle stiffness and inter-particle sliding friction in milling of particles," *Particuology*, vol. 16, pp. 54–59, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.partic.2014.04.003

A comparative study of Minimum Variance Distortionless Response beamforming, Linearly Constrained Minimum Variance beamforming and metaheuristic solved null-steering beamforming

Mr. Robert Macharia Maina and Dr. Kibet Lang'at and Dr. P. K. Kihato

Abstract—Digital beamforming is of paramount significance in addressing wireless communication systems' capacity enhancement. Among well established digital beamforming mechanisms are Minimum Variance Distortionless Response (MVDR) and Linearly Constrained Minimum Variance (LCMV) techniques. It would be worthwhile to establish the relative performance of the afore-mentioned techniques in beamforming in various situations. In this paper, MVDR and LCMV techniques performance comparison is addressed from the perspective of narrowband beamforming (in reception mode), with the variation parameter being signal/ interferer spatial separation. The LCMV beamformer has been constrained to yield nulls in interferer Direction of Arrivals (DoAs). Also studied is a Particle Swarm Optimization (PSO) solved null-steering beamformer. It is established that the PSO solved null-steering beamformer outperforms the LCMV and MVDR techniques. MATLAB software has been used as the simulation tool.

Keywords—Digital beamforming, Linearly Constrained Minimum Variance, Minimum Variance Distortionless Response, Null steering

I. INTRODUCTION

Beamforming in wireless communication systems is of paramount significance towards achieving capacity enhancement. Beamforming is essentially a spatial filtering action [1]. Among techniques utilized in digital beamforming are Minimum Variance Distortionless Response (MVDR) and Linearly Constrained Minimum Variance (LCMV) methods. Use of metaheuristics in beamforming procedures such as beam steering, null steering, reference signal aided beamforming among others is an active research area. Beamforming on the basis of MVDR, LCMV and null steering techniques is a focus of the work presented in this paper.

Essentially, a reception digital beamformer operates on the basis of weighting signals received at elements of a suitably designed array. Taking into consideration an M element antenna array, a digital reception beamformer output can be framed as per (1). Reference can be made to [1].

$$y = \bar{w}^H \bar{x} \tag{1}$$

In (1):

- y is the beamformer output.
- $\bar{w}$ is the beamformer weights vector (of size M).
- $\bar{x}$ is an M sized vector containing signals observed at the array elements.

Mr. Robert Macharia Maina, Department of Telecommunication and Information Engineering, JKUAT (e-mail: robertisaacm@gmail.com).

Dr. Kibet Lang'at, Department of Telecommunication and Information Engineering, JKUAT.

Dr. P. K. Kihato, Department of Electrical and Electronic Engineering, JKUAT.

II. MINIMUM VARIANCE DISTORTIONLESS RESPONSE BEAMFORMER

The MVDR beamformer is designed with an aim of minimizing undesired signal power (interference and noise) whilst simultaneously maintaining maximal response in the Direction of Arrival (DoA) of the desired signal. This can be expressed mathematically as per (2).

$$\text{Minimize} \qquad \bar{w}^H \bar{R}_{uu} \bar{w} \tag{2a}$$

$$\text{Subject to} \qquad \bar{w}^H \cdot \bar{a}_s = 1 \tag{2b}$$

In (2), R_{uu} is the undesired signal correlation matrix and $\bar{a}_s$ is the array response vector corresponding to the desired signal DoA.

Studies involving the MVDR beamforming technique can be found in [2]–[5].

III. LINEARLY CONSTRAINED MINIMUM VARIANCE BEAMFORMER

The LCMV beamformer is designed with an aim of minimizing the total output power of an array subject to some constraints: generate maximal gain in the desired signal DoA and generate minimal gain in the undesired signals DoAs.

Studies involving the LCMV beamforming technique can be found in [6]–[9].

IV. NULL STEERING BEAMFORMER

The null steering beamformer is designed with an aim of generating an array response with maximum gain the desired signal DoA and nulls in the undesired (interferer) signal DoAs.

(3a) gives the array response magnitude (AF_s) corresponding to the desired signal DoA. (3b) gives the array response magnitude (AF_n) corresponding to an interfering signal DoA. The null steering condition to be optimized can be expressed as per (4).

$$AF_s = \bar{w}^H \cdot \bar{a}(\theta_s) \tag{3a}$$

$$AF_n = \bar{w}^H \cdot \bar{a}(\theta_n) \tag{3b}$$

$$J(\bar{w}) = AF_s - AF_n = \bar{w}^H \cdot \bar{a}(\theta_s) - \bar{w}^H \cdot \bar{a}(\theta_n) \quad (4)$$

$J(\bar{w})$ in (4) can be optimized using a metaheuristic algorithm, an action carried out in this paper.

Studies involving the null steering beamforming technique can be found in [10], [11].

V. METHODOLOGY

The problem at hand is essentially a reception beamforming problem. A receive linear antenna array with 10 isotropic elements is utilized. The desired signal is framed as having a DoA of 0 degrees azimuth, 0 degrees elevation. Two interferers are implemented, whose DoAs are framed as per Table I.

TABLE I
INTERFERER DOAS.

	Interferer 1 DoA		Interferer 2 DoA	
	Azimuth	Elevation	Azimuth	Elevation
Trial 1	-30	0	30	0
Trial 2	-20	0	20	0
Trial 3	-10	0	10	0
Trial 4	-5	0	5	0

The desired signal is framed as a sinusoid as per Fig. 1.

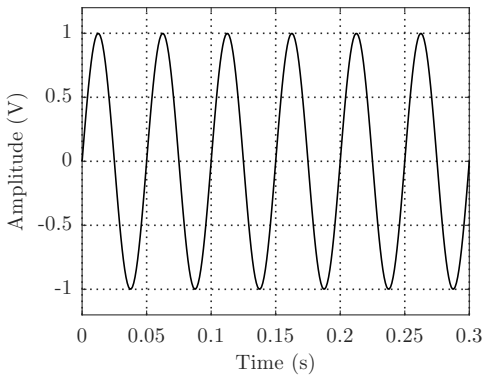


Fig. 1. Desired signal.

The interferes are framed as random signals as per Figs. 2 and 3.

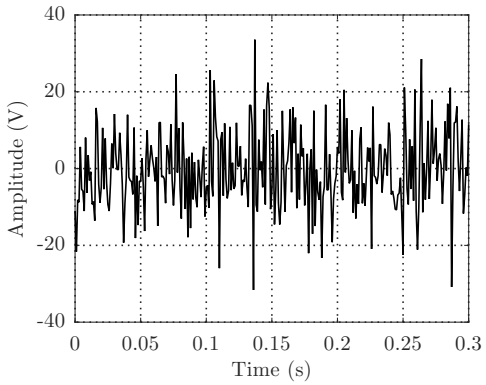


Fig. 2. Interferer 1.

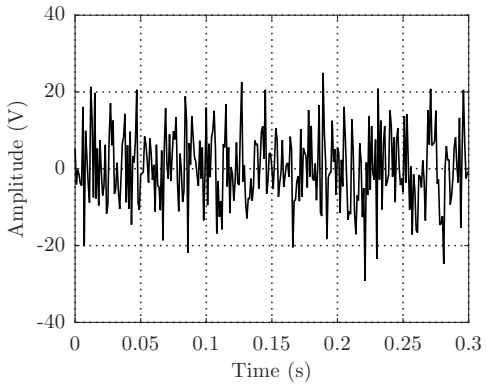


Fig. 3. Interferer 2.

A noisy channel resulting in a Signal to Noise Ratio (SNR) figure of 50dB is implemented. MVDR, LCMV, and Particle Swarm Optimization (PSO) solved null-steering beamforming schemes are utilized in solving the beamforming problem at hand. The LCMV beamformer has been constrained to yield nulls in interferer DoAs. A comparative study among the beamforming schemes is carried out with angular interferer separation as the study variable.

VI. RESULTS

The comparative study results are herein presented. The first trial results are comprehensively presented (including graphical presentations of signals received at each array element). For reasons of brevity, the other trial results are concisely presented.

A. Trial 1: 60 degrees interferer separation

The signals observed at the array elements are as per Fig. 4 to Fig. 13.

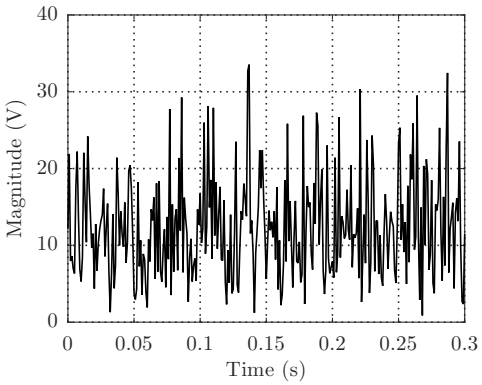


Fig. 4. Signal observed at array element 1.

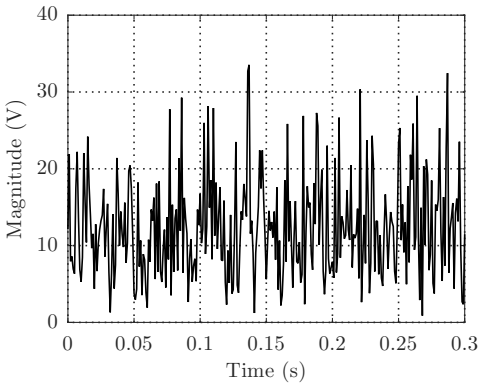


Fig. 5. Signal observed at array element 2.

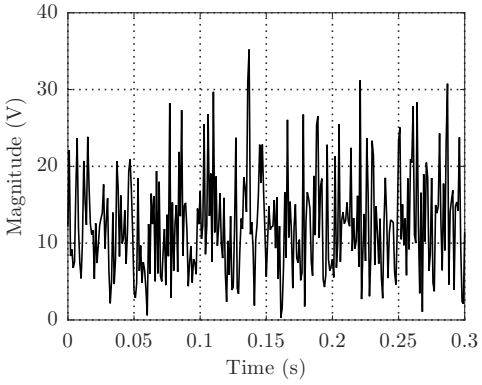


Fig. 6. Signal observed at array element 3.

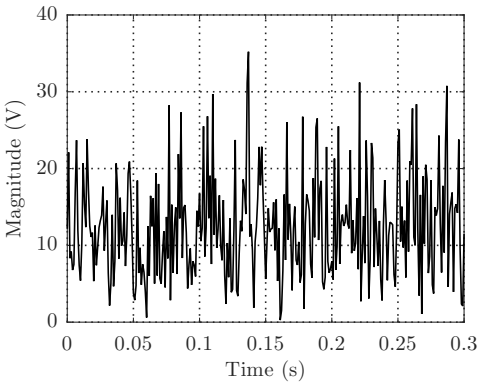


Fig. 7. Signal observed at array element 4.

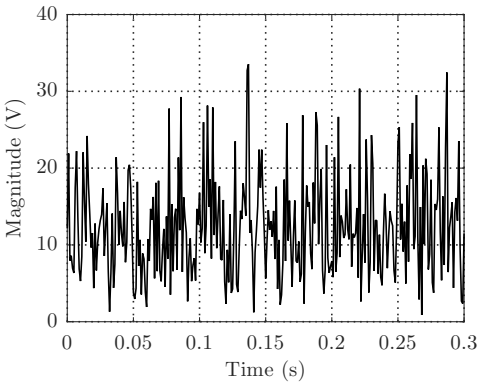


Fig. 8. Signal observed at array element 5.

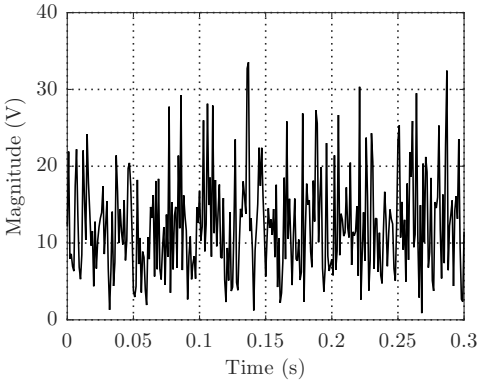


Fig. 9. Signal observed at array element 6.

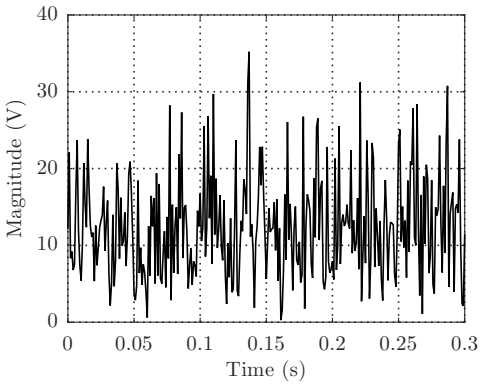


Fig. 10. Signal observed at array element 7.

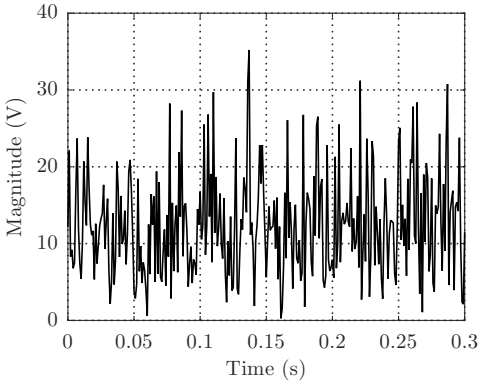


Fig. 11. Signal observed at array element 8.

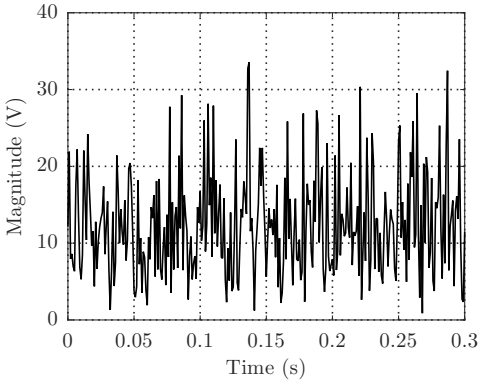


Fig. 12. Signal observed at array element 9.

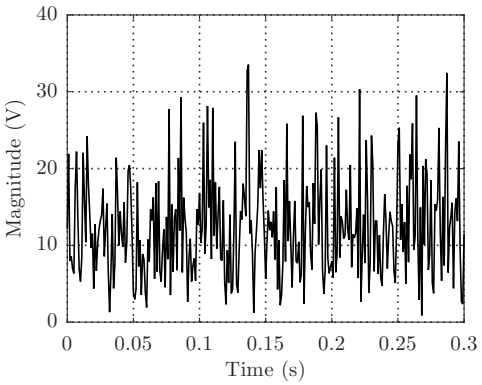


Fig. 13. Signal observed at array element 10.

Upon beamforming using the MVDR beamformer, the signal observed at the beamformer output is as per Fig. 14.

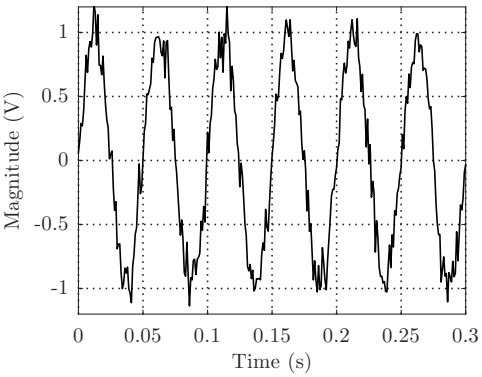


Fig. 14. Signal observed at beamformer output upon using the MVDR beamformer.

The MVDR beamformer resultant array response pattern is as per Fig. 15.

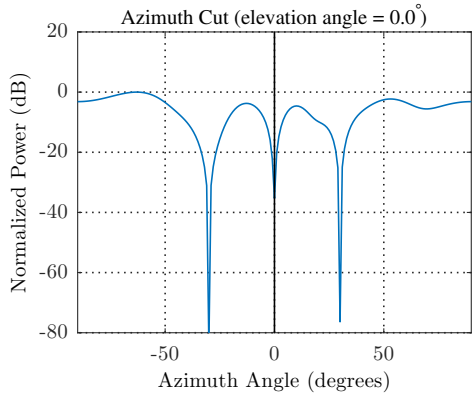


Fig. 15. Beamformer array response upon using the MVDR beamformer.

Upon beamforming using the LCMV beamformer, the signal observed at the beamformer output is as per Fig. 16.

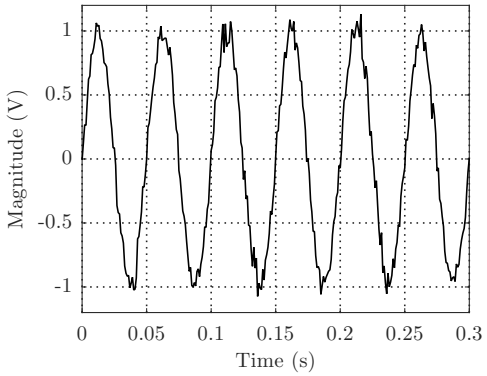


Fig. 16. Signal observed at beamformer output upon using the LCMV beamformer.

The LCMV beamformer resultant array response pattern is as per Fig. 17.

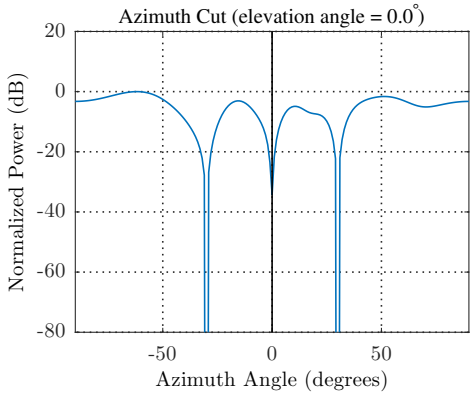


Fig. 17. Beamformer array response upon using the LCMV beamformer.

Upon beamforming using the PSO beamformer, the signal observed at the beamformer output is as per Fig. 18.

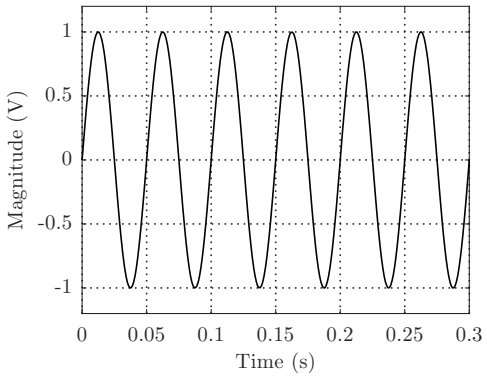


Fig. 18. Signal observed at beamformer output upon using the PSO beamformer.

The PSO beamformer resultant array response pattern is as per Fig. 19.

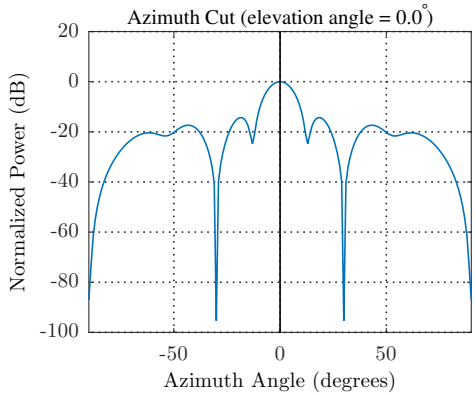


Fig. 19. Beamformer array response upon using the PSO beamformer.

Fig. 20 presents a comparative array response plot.

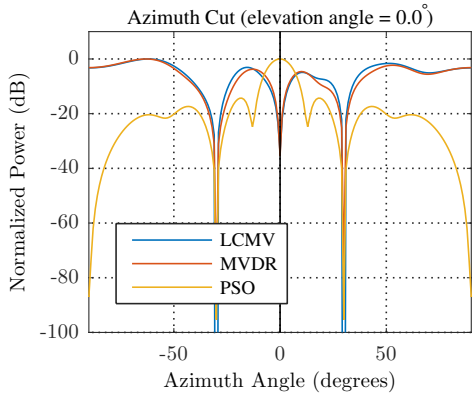


Fig. 20. Comparative array response plot.

From Fig. 20, it is easy to note the superior performance of the PSO solved null-steering beamformer over the MVDR and LCMV beamformers.

B. Trial 2: 40 degrees interferer separation

Upon beamforming using the MVDR beamformer, the signal observed at the beamformer output is as per Fig. 21.

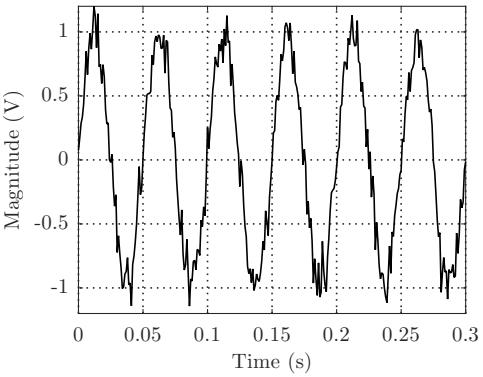


Fig. 21. Signal observed at beamformer output upon using the MVDR beamformer.

The MVDR beamformer resultant array response pattern is as per Fig. 22.

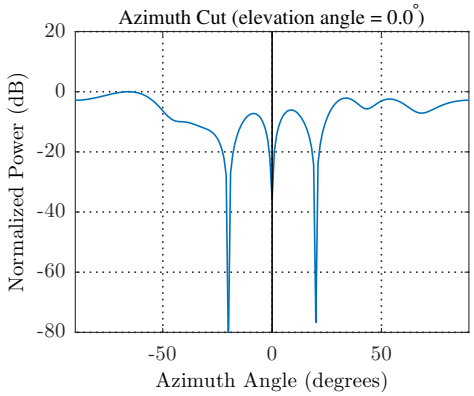


Fig. 22. Beamformer array response upon using the MVDR beamformer.

Upon beamforming using the LCMV beamformer, the signal observed at the beamformer output is as per Fig. 23.

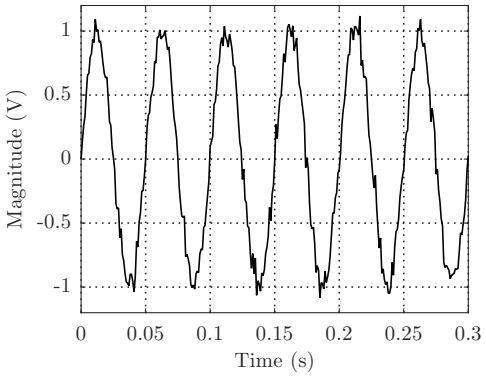


Fig. 23. Signal observed at beamformer output upon using the LCMV beamformer.

The LCMV beamformer resultant array response pattern is as per Fig. 24.

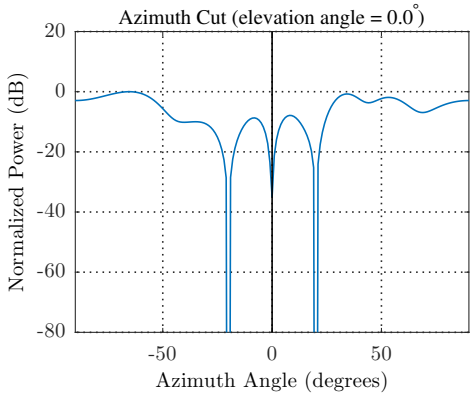


Fig. 24. Beamformer array response upon using the LCMV beamformer.

Upon beamforming using the PSO beamformer, the signal observed at the beamformer output is as per Fig. 25.

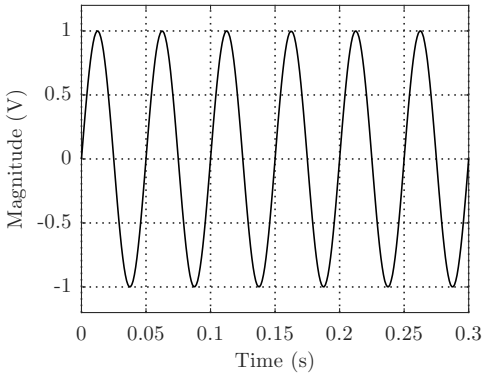


Fig. 25. Signal observed at beamformer output upon using the PSO beamformer.

The PSO beamformer resultant array response pattern is as per Fig. 26.

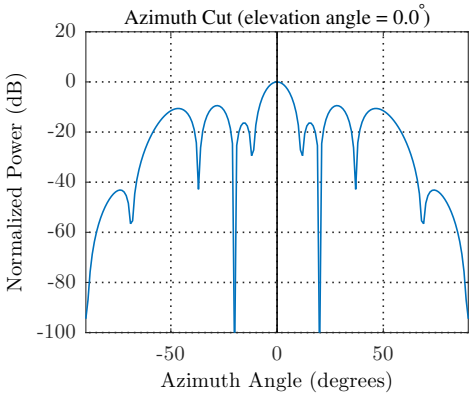


Fig. 26. Beamformer array response upon using the PSO beamformer.

Fig. 27 presents a comparative array response plot.

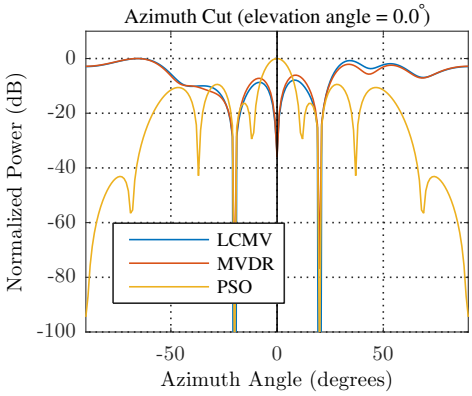


Fig. 27. Comparative array response plot.

From Fig. 27, it is easy to note the superior performance of the PSO solved null-steering beamformer over the MVDR and LCMV beamformers.

C. Trial 3: 20 degrees interferer separation

Upon beamforming using the MVDR beamformer, the signal observed at the beamformer output is as per Fig. 28.

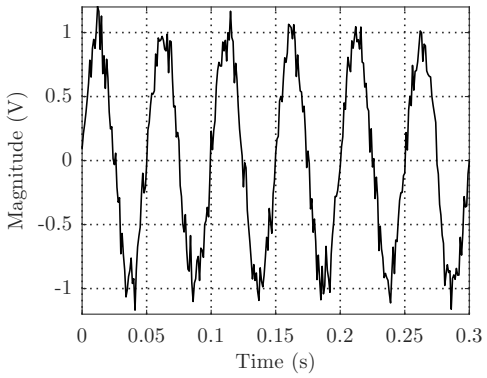


Fig. 28. Signal observed at beamformer output upon using the MVDR beamformer.

The MVDR beamformer resultant array response pattern is as per Fig. 29.

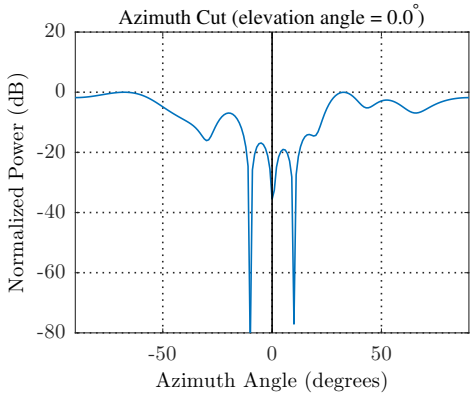


Fig. 29. Beamformer array response upon using the MVDR beamformer.

Upon beamforming using the LCMV beamformer, the signal observed at the beamformer output is as per Fig. 30.

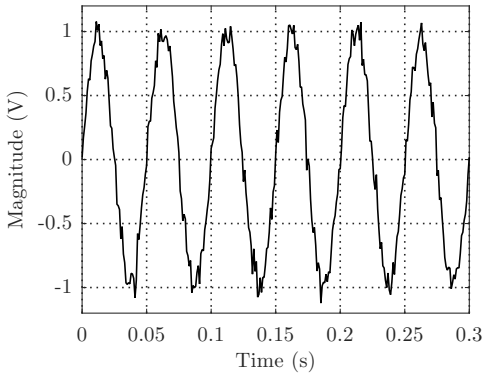


Fig. 30. Signal observed at beamformer output upon using the LCMV beamformer.

The LCMV beamformer resultant array response pattern is as per Fig. 31.

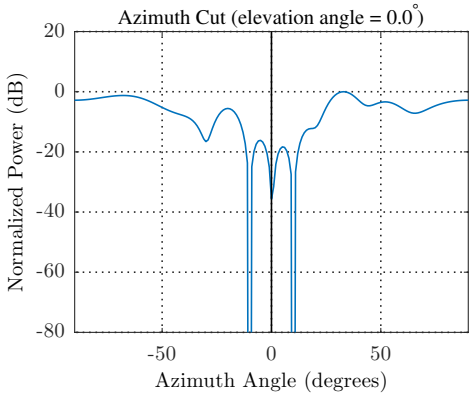


Fig. 31. Beamformer array response upon using the LCMV beamformer.

Upon beamforming using the PSO beamformer, the signal observed at the beamformer output is as per Fig. 32.

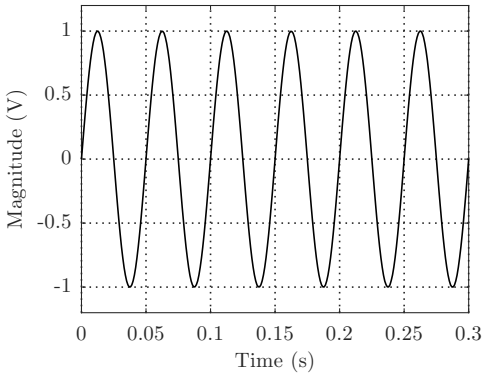


Fig. 32. Signal observed at beamformer output upon using the PSO beamformer.

The PSO beamformer resultant array response pattern is as per Fig. 33.

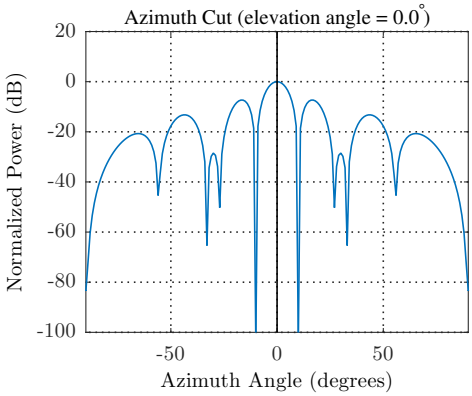


Fig. 33. Beamformer array response upon using the PSO beamformer.

Fig. 34 presents a comparative array response plot.

From Fig. 34, it is easy to note the superior performance of the PSO solved null-steering beamformer over the MVDR and LCMV beamformers.

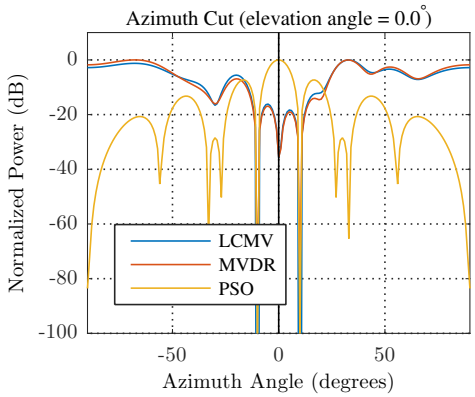


Fig. 34. Comparative array response plot.

D. Trial 4: 10 degrees interferer separation

Upon beamforming using the MVDR beamformer, the signal observed at the beamformer output is as per Fig. 35.

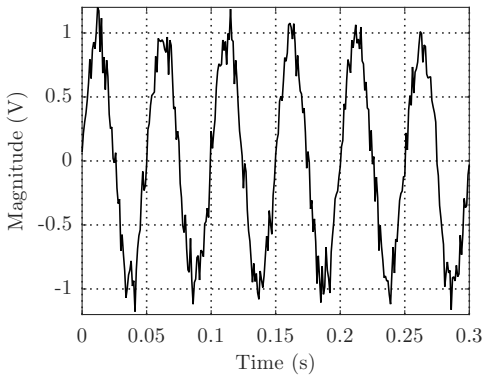


Fig. 35. Signal observed at beamformer output upon using the MVDR beamformer.

The MVDR beamformer resultant array response pattern is as per Fig. 36.

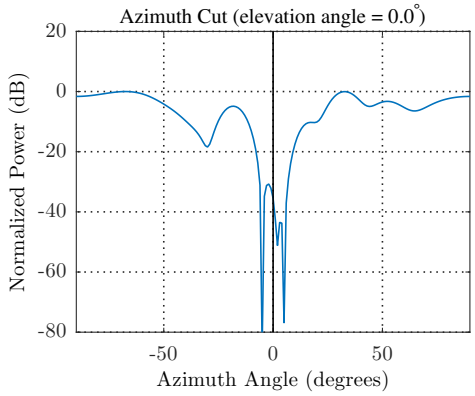


Fig. 36. Beamformer array response upon using the MVDR beamformer.

Upon beamforming using the LCMV beamformer, the signal observed at the beamformer output is as per Fig. 37.

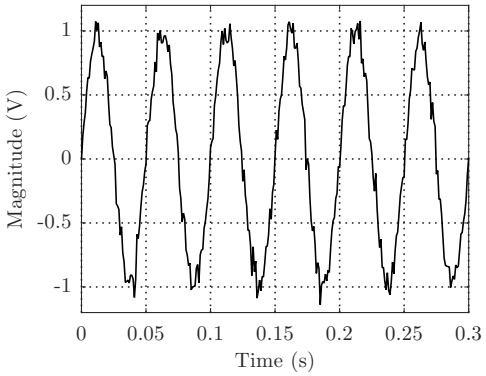


Fig. 37. Signal observed at beamformer output upon using the LCMV beamformer.

The LCMV beamformer resultant array response pattern is as per Fig. 38.

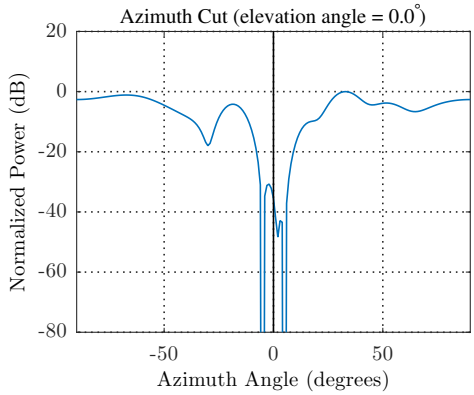


Fig. 38. Beamformer array response upon using the LCMV beamformer.

Upon beamforming using the PSO beamformer, the signal observed at the beamformer output is as per Fig. 39.

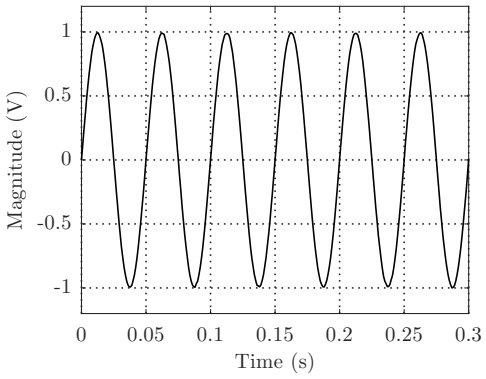


Fig. 39. Signal observed at beamformer output upon using the PSO beamformer.

The PSO beamformer resultant array response pattern is as per Fig. 40.

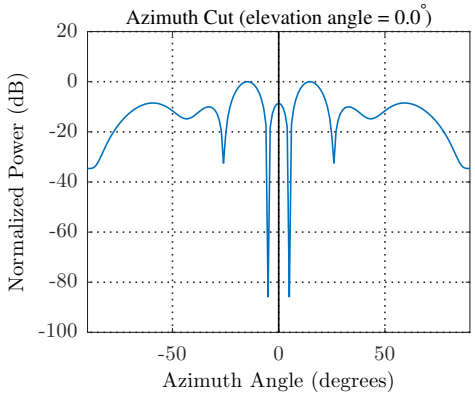


Fig. 40. Beamformer array response upon using the PSO beamformer.

Fig. 41 presents a comparative array response plot.

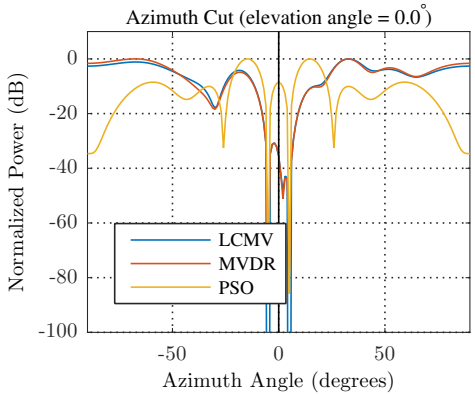


Fig. 41. Comparative array response plot.

From Fig. 41, it is easy to note the superior performance of the PSO solved null-steering beamformer over the MVDR and LCMV beamformers.

E. Overall SINR comparison

Going by Table II, the superior performance of the PSO solved null steering beamformer is evident. The beamformer yields good results even with narrow interferer angular separation. The LCMV beamformer outperforms the MVDR beamformer in all trials.

TABLE II
SINR RESULTS IN DB.

	Trial 1	Trial 2	Trial 3	Trial 4
LCMV Mean	17.643	17.213	16.932	16.789
LCMV Min	9.838	9.330	8.388	8.317
MVDR Mean	14.838	14.522	14.633	14.508
MVDR Min	6.299	5.874	6.872	7.188
PSO Mean	55.284	56.103	56.045	40.420
PSO Min	47.907	48.922	48.500	32.656

VII. CONCLUSION

In this paper, MVDR and LCMV techniques performance comparison has been addressed from the perspective of narrowband beamforming (in reception mode), with the variation parameter being signal/ interferer spatial separation. The LCMV beamformer has been constrained to yield nulls in interferer DoAs. Also studied is a PSO solved null-steering beamformer. From MATLAB software simulations, it has been established that the PSO solved null-steering beamformer outperforms the LCMV and MVDR techniques.

REFERENCES

[1] F. B. Gross, *Smart Antennas for Wireless Communications*. New York: McGraw-Hill, 2nd ed., 2015.
[2] B. Cauchi, I. Kodrasi, R. Rehr, S. Gerlach, A. Jukić, T. Gerkmann, S. Doclo, and S. Goetze, "Combination of mvdr beamforming and single-channel spectral processing for enhancing noisy and reverberant speech," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, p. 61, Jul 2015.
[3] R. Qian, M. Sellathurai, and D. Wilcox, "A study on mvdr beamforming applied to an espar antenna," *IEEE Signal Processing Letters*, vol. 22, pp. 67–70, Jan 2015.
[4] S. Araki, M. Okada, T. Higuchi, A. Ogawa, and T. Nakatani, "Spatial correlation model based observation vector clustering and mvdr beamforming for meeting recognition," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 385–389, March 2016.
[5] M. O'Connor, W. B. Kleijn, and T. Abhayapala, "Distributed sparse mvdr beamforming using the bi-alternating direction method of multipliers," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 106–110, March 2016.
[6] C. Doroody, S. K. Tiong, and S. Darzi, "Sinr improvement using firefly algorithm (fa) for linear constrained minimum variance (lcmv) beamforming technique," in *2015 International Conference on Computer, Communications, and Control Technology (I4CT)*, pp. 441–445, April 2015.
[7] T. Sherson, W. B. Kleijn, and R. Heusdens, "A distributed algorithm for robust lcmv beamforming," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 101–105, March 2016.
[8] A. Hassani, A. Bertrand, and M. Moonen, "Lcmv beamforming with subspace projection for multi-speaker speech enhancement," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 91–95, March 2016.
[9] A. Hassani, J. Plata-Chaves, M. H. Bahari, M. Moonen, and A. Bertrand, "Multi-task wireless sensor network for joint distributed node-specific signal enhancement, lcmv beamforming and doa estimation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, pp. 518–533, April 2017.
[10] S. Saeed, I. M. Qureshi, A. Basit, A. Salman, and W. Khan, "Cognitive null steering in frequency diverse array radars," *International Journal of Microwave and Wireless Technologies*, vol. 9, no. 1, p. 2533, 2017.
[11] S. N. Monteiro and H.G.Virani, "Improved null steering with sidelobe canceller for linear antenna arrays," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 4, pp. 97–100, March 2015.

Wi-Fi Signal Indoor LOS Coverage modeling using PSO-ANFIS

Omae M. O, Ndungu E. N and Kibet P. L.

***Abstract*-**Wireless local area networks (WLANS) are becoming very popular in our daily communications applications. Currently in kenya a number of internet service providers like Safaricom, Zuku and others are providing internet access using Wi-Fi. I believe this is replicated world over. This has necessitated studies on these systems to improve on the quality of service (QoS) provided to the users. Different methods have been used in signal modeling. This study is aimed at predicting Wi-Fi signal propagation along a corridor using Particle Swarm Optimization (PSO) trained Adaptive Neural Fuzzy Inference System (ANFIS). The root mean square and standard deviation of the predicted signal were determined. The study was undertaken using a Wi-Fi router as the transmitter and a mobile phone as the receiver in the process of data collection. The measured values were then used in PSO-ANFIS modeling. It was found that the predicted values were close to actual measured values as from the undertaken analysis.
Keywords; Wi-Fi, QoS, WLANS, ANFIS

I. INTRODUCTION

Wi-Fi networks form one of the largest market segments of wireless networks. Coverage in line of sight (LOS) environments is limited both by physical obstacles and structural barriers, while in built environments, the main obstacles are walls [1]. What is common for both is interference in the wireless spectrum. The most commonly used ISM bands for Wi-Fi networks are at 2.4 GHz and 5 GHz, and the signals at such high frequencies do not easily pass through the obstacles. To increase connectivity and extend coverage, Wi-Fi networks use limited transmission powers, typically up to 100 mW. This gives connectivity of a few tens of meters, even through walls. At the same time, line-of-sight connectivity may reach significantly greater distances, causing far away nodes to interfere in very unusual patterns.

ANFIS is one of the most current techniques used in function approximation besides other very many applications like classification. The technique is obtained by combining the Neural Networks and Fuzzy Logic concepts which are based on numerical analysis and natural language respectively [3].

PSO originally by Doctor Kennedy and E berhart in 1995, used to train ANFIS and other AI processes is based on the intelligence of swarms as they move in search of food [9].

This study investigated the prediction of signal coverage of Wi-Fi networks using PSO trained ANFIS.

A. Statement of the problem

WLANS are currently becoming a very important concept in our lives. Scientists have done various studies in regard to this technology and continue to do the same to ensure quality of service (QoS) is improved to the ever growing number of users. In view of this the idea of also adding to the progressing research in this field led to the study of prediction of Wi-Fi signal using PSO-ANFIS which is commonly used in approximating functions because of its advantages that include high accuracy and better computational efficiency.

B. Research objectives

Main objective;

To predict Wi-Fi signal coverage using PSO trained ANFIS.

Specific objectives

1. Measure signal strength with variation of distance along a corridor.
2. Use PSO trained ANFIS to model the measured signal.

II. LITERATURE REVIEW

A. Introduction

Wireless networking works by sending radio transmissions on specific frequencies where listening devices can receive them. The necessary radio transmitters and receivers are built into Wi-Fi enabled equipment like routers, laptops and phones. Antennas are also key components of these radio communication systems, picking up incoming signals or radiating outgoing Wi-Fi signals. Some Wi-Fi antennas, particularly on routers, may be mounted externally while others are embedded inside the device's hardware enclosure [2].

ANFIS combines the advantages of both neural network and fuzzy logic in its operation resulting to a powerful tool in approximating functions [3].

PSO finds the optimal solution by simulating the social behavior of groups as fish schooling or bird flocking. A group can achieve the objective effectively by using the common information of every particle (global), and the information owned by the particle itself (personal) [9].

B. Effect of distance

Signal attenuation over distance is observed when the mean received signal power is attenuated as a function of the distance from the transmitter. The most common form of this is often called free space loss and is due to the signal power being

Omae M. O, Ndungu E. N, Kibet P. L.School of Electrical, Electronic and Information Engineering, Department of Telecommunication & Information Engineering, JKUAT (phone: +2540722805012;+2540721366349, +2540724833749 fax: +2546752711; e-mail: m_oteri@yahoo.com,ndunguen@yahoo.com, kibetlp@yahoo.com).

spread out over the surface area of an increasing sphere as the receiver moves farther from the transmitter.

In addition to free space loss effects, the signal experiences decay due to ground wave loss although this typically only comes into play for very large distances (on the order of kilometers) [].

C. Multipath Propagation

Multipath results from the fact that the propagation channel consists of several obstacles and reflectors. Thus, the received signal arrives as an unpredictable set of reflections and/or direct waves each with its own degree of attenuation and delay. The delay spread is a parameter commonly used to quantify multipath effects. Multipath leads to variations in the received signal strength over frequency and antenna location [].

D. Rate of fading

Time variation of the channel occurs if the communicating device (antenna) and components of its environment are in motion. Closely related to Doppler shifting, time variation in conjunction with multipath transmission leads to variation of the instantaneous received signal strength about the mean power level as the receiver moves over distances on the order of less than a single carrier wavelength.

The degree of time variation in an outdoor system is much more than that of an indoor system. One manifestation of time variation is as spreading in the frequency domain (Doppler spreading). The frequency in our case varied from 2412 to 2467 MHz.

E. Free space path loss

Free space path loss (FSPL) is the loss in signal strength that occurs when an electromagnetic wave travels over a line of sight (LOS) path in free space. In such a circumstance, there are no obstacles that might cause the signal to be reflected, refracted or that might cause additional attenuation. [4]

When calculating thus, factors relating to the transmitter power, antenna gains or the receiver sensitivity levels are not considered and only the loss along the path itself is considered. As a signal moves away from the transmitter, it keeps spreading out in the form of a sphere increasing the sphere's surface area with increase in distance thus, the intensity of the signal decreases. It can be deduced that the signal decreases in a manner that is inversely proportional to the square of the distance from the source of the radio signal in free space.

Losses are experienced in radio wave communication links as the signal is sent from the source to the destination. One type of such losses is path losses. These occur due to effects along the transmission media. Under path losses we have free space losses among others [5]. These are highly affected by variation of distance and frequency.

The received power at the destination in dB is given by:-

$$P_R = P_t G_t G_R/(4\pi d/\lambda)^2 \quad (1)$$

$$P_{R} = P_{t\,dB} + G_{t\,dB} + G_{R\,dB} - FSL_{\,dB} \quad (2)$$

P_R is received power

P_t is the transmitted power

G_t is the transmitter gain

G_R is receiver gain

This is referred to as Friis equation which is the link equation.

Most RF comparisons and measurements are performed in decibels. This gives an easy and consistent method to compare the signal levels present at various points. Accordingly it is very convenient to express the free space path loss formula, FSPL, in terms of decibels. It is easy to take the basic free space path loss equation and manipulate into a form that can be expressed in a logarithmic format.

Free space losses (FSL) is given by:-

$$FSL = 32.44 + 20\log d + 20\log f \quad (3)$$

Where;

FSL= free space losses in dB

d= distance between the source and destination in kms

f= frequency

In this work, the apparatus used have the following specifications:

Mobile Phone Receiver

Tecno R7 with G(r) as +4dB was used.

D-Link DIR 605L router (Transmitter)

P (t) = +15dBm; G (t) = +4dBi

Therefore, P (D-Link) = 15 + 4 + 4 = +23dB.

The fundamental design of and plan of indoor wireless network depends on the measurement and analysis of the Wi-Fi signal. Distance is one of the major contributors of the attenuation of the radio signal propagation known as the path loss [6]. The signal received by the user reduces in power with the distance it traverses following an inverse square law. For an ideal condition the power of the signal is given by

$P_R = P_{t\,dB} + G_{t\,dB} + G_{R\,dB} - FSL_{\,dB}$

Where P_R is the power transmitted

P_t is power of the router

G_t is the gain of the router

G_R is the antennae gain for the mobile device/laptop

FSL is given by $32.44 + 20logd + 20logf$ where d is the distance in km and f is the frequency in MHz [7].

F. Other losses

Apart from free spaces there are other losses that contribute to the reduction of the received signal strength that include atmospheric losses, antenna misalignment, polarization mismatch and losses due a number of path obstacles like walls, people and furniture [].

G. Adaptive Neuro-Fuzzy Inference System (ANFIS)

Adaptive Neuro-Fuzzy Inference System (ANFIS) otherwise referred to as Adaptive Network-based Fuzzy Inference System was proposed in [7]. ANFIS is a blend of Fuzzy Logic (FL) and Artificial Neural Network (ANN) that captures the strengths and offsets the limitations of both techniques for building Inference Systems (IS) with improved results and enhanced intelligence. Fuzzy logic is associated with the theory of fuzzy set, which relates to classes of objects with rough boundaries in which membership is a matter of degree. It is an extensive multivalued logical system that departs in concept and substance from the traditional multivalued logical systems. Much of fuzzy logic may be viewed as a platform for computing

with words rather than numbers. The use of words for computing is closer to human intuition and exploits the tolerance for imprecision, thereby lowering the cost of the solution [8]. However, there are no known appropriate or well-established methods of defining rules and membership functions based on human knowledge and experience. Artificial Neural Networks are made up of simple processing elements operating concurrently. These elements model the biological nervous system, with the network functions predominantly determined by the connections between the elements. Neural Networks have the ability to learn from data by adjusting the values of the connections (weights) between the elements. Merging these two artificial intelligence paradigms together offers the learning power of neural networks and the knowledge representation of fuzzy logic for making inferences from observations.

H. Basic ANFIS Architecture

The ANFIS architecture described here is based on type 3 fuzzy inference system (other popular types are the type 1 and type 2). In the type 3 inference system, the Takagi and Sugeno's (TKS) if-then rules are used [3]. The output of each rule is obtained by adding a constant term to the linear combination of the input variables. Final output is then computed by taking the weighted average of each rule's output. The type 3 ANFIS architecture with two inputs (x and y) and one output, z, is shown in Fig. 1.

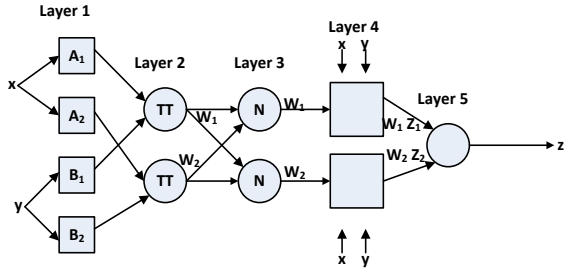


Fig. 1. Type 3 ANFIS Architecture.

$Rule\ 1: If\ x\ is\ A_1\ and\ y\ is\ B_1, then\ z_1 = p_1x + q_1y + r_1$
$Rule\ 2: If\ x\ is\ A_2\ and\ y\ is\ B_2, then\ z_2 = p_2x + q_2y + r_2$

The ANFIS structure is the functional equivalent of a supervised, feed-forward neural network with one input layer, three hidden layers and one output layer, whose functionality are as described below:

Layer 1 (Fuzzy Layer): Every node in this layer is an adaptive layer that generates the membership grades of the input vectors. Usually, a bell-shaped (Gaussian) function with maximum equal to 1 and minimum equal to 0 is used for implementing the node function:

$$O_i^1 = f(x, a, b, c) = \mu_{Ai}(x) = \frac{1}{1+|(x-c_i)/a_i|^{2b_i}}$$

$$\mu_{Ai}(x) = \exp\{-\left[\left(\frac{x-c_i}{a_i}\right)^2\right]^{b_i}\} \quad (4)$$

Where O_i^1 is the output of the i^{th}node in the first layer, $\mu_{Ai}(x)$ is the membership function of input in the linguistic variable A_i. The parameter set $\{a_i, b_i, c_i\}$ are responsible for defining the shapes of the membership functions. These parameters are called premise parameters.

Layer 2 (Product Layer): Each mode in this layer determines the firing strength of a rule by multiplying the membership functions associated with the rules. The nodes in this layer are fixed in nature. The firing strength of a particular rule (the output of a node) is given by:

$$w_i = O_i^2 = \mu_{Ai}(x).\mu_{Bi}(y), i = 1, 2 \quad (5)$$

Any other T-norm operator that performs fuzzy AND operation can be used in this layer.

Layer 3 (Normalized Layer): This layer consists of fixed nodes that are used to compute the ratio of the i[th] rule's firing strength to the total of all firing strengths:

$$\overline{w} = O_i^3 = \frac{w_i}{w_1+w_2}, i = 1, 2, \quad (6)$$

The outputs of this layer are otherwise known as normalized firing strength for convenience.

Layer 4 (Defuzzify Layer): This is an adaptive layer with node function given by:

$$\overline{w_i}z_i = O_i^4 = \overline{w_i}(p_ix + q_iy + r_i) \quad (7)$$

This layer essentially computes the contribution of each rule to the overall output. It is defuzzification layer and provides output values resulting from the inference of rules. The parameters in this layer $\{p_i,\ q_i,\ r_i\}$ are known as consequent parameters.

Layer 5 (Total Output Layer): There is only one fixed node in this layer. It computes the overall output as the summation of contribution from each rule:

$$\sum_i \overline{w_i}z_i = O_i^5 = \sum_i \frac{w_iz_i}{\sum_i w_i} \quad (8)$$

I. Particle Swarm Optimization (PSO)

PSO is a global optimization technique that was developed by Eberhart and Kennedy in 1995 [12], the underlying motivation of PSO algorithm was the social behavior observable in nature, such as flocks of birds and schools of fish in order to guide swarms of particles towards the most promising regions of the search space. PSO exhibits a good performance in finding solutions to static optimization problems where it is considered to be better than other algorithms like Genetic Algorithm [14]. It exploits a population of individuals to synchronously probe promising regions of the search space. In this context, the population is called a swarm and the individuals (i.e. the search points) are referred to as particles. Each particle in the swarm represents a candidate solution to the optimization problem. In

a PSO system, each particle moves with an adaptable velocity through the search space, adjusting its position in the search space according to own experience and that of neighboring particles, then it retains a memory of the best position it ever encountered, a particle therefore makes use of the best position encountered by itself and the best position of neighbors to position itself towards the global minimum. The effect is that particles "fly" towards the global minimum, while still searching a wide area around the best solution [11]. The performance of each particle (i.e. the "closeness" of a particle to the global minimum) is measured according to a predefined fitness function which is related to the problem being solved. For the purposes of this research, a particle represents the weight vector of NNs, including biases. The dimension of the search space is therefore the total number of weights and biases [11].

The iterative approach of PSO can be described by the following steps:

Step 1: Initialize a population size, positions and velocities of agents, and the number of weights and biases.

Step 2: The current best fitness achieved by particle *p* is set as *pbest*. The *pbest* with best value is set as *gbest* and this value is stored.

Step 3: Evaluate the desired optimization fitness function f_p for each particle as the Mean Square Error (MSE) over a given data set.

Step 4: Compare the evaluated fitness value f_p of each particle with its *pbest* value. If $f_p <$ pbest then pbest = f_p and $best_{xp} = x_p$, x_p is the current coordinates of particle p, and bestxp is the coordinates corresponding to particle *p's* best fitness so far.

Step 5: The objective function value is calculated for new positions of each particle. If a better position is achieved by an agent, *pbest* value is replaced by the current value. As in Step 1, *gbest* value is selected among *pbest* values. If the new *gbest* value is better than previous *gbest* value, the *gbest* value is replaced by the current *gbest* value and this value is stored. If $f_p <$ gbest then gbest = p, where *gbest* is the particle having the overall best fitness over all particles in the swarm.

Step 6: Change the velocity and location of the particle according to Equations 9 and 10, respectively.

Step 7: Fly each particle p according to Equation 9.

Step 8: If the maximum number of predetermined iterations (epochs) is exceeded, then stop; otherwise Loop to step 3 until convergence.

$$V_i = wV_{i-1} + acc * rand() * \left(best_{xp} - xp\right) + acc * rand() * \left(best_{xgbest} - xp\right) \quad (9)$$

Where *acc* is the acceleration constant that controls how far particles fly from one another, and *rand* returns a uniform random number between 0 and 1.

$$xp = xpp + V_i \quad (10)$$

V_i is the current velocity, V_{i-1} is the previous velocity, xp is the present location of the particle, xpp is the previous location of the particle, and *i* is the particle index. In step 5 the coordinates bestxp and bestxgbest are used to pull the particles towards the global minimum [11].

Learning by PSO

To develop an accurate process model using ANFIS, the training, and validation processes are among the important steps. In the training process, a set of input-output patterns is repeated to the ANFIS. From that, weights of all the interconnections between neurons are adjusted until the specified input yields the desired output. Through these activities, the ANFIS learns the correct input-output response behavior [11].

The way PSO will be employed for updating the ANFIS parameters is explained in this section. The ANFIS has two types of parameters which need training, the antecedent part parameters and the conclusion part parameters. The membership functions are assumed Gaussian as in equation (3.4), and their parameters are $\{a_i, b_i, c_i\}$, where a_i is the variance of membership functions and c_i is the center of membership functions (MFs). Also is b_i a trainable parameter. The parameters of conclusion part are trained and here are represented with $\{p_i,\ q_i,\ r_i\}$ [11].

Applying PSO for Training ANFIS parameters

There are 3 sets of trainable parameters in antecedent part $\{a_i, b_i, c_i\}$, each of these parameters has N genes. Where, N represents the number of MFs. The conclusion parts parameters ($\{p_i,\ q_i,\ r_i\}$) also are trained during optimization algorithm. Each chromosome in conclusion part has $(I+1)\times R$ genes that R is equal to Number of rules and *I* denotes dimension of data inputs. For example each chromosome in conclusion part in figure3 has 6 genes. The fitness is defined as mean square error (MSE) [11].

Parameters are initialized randomly in first step and then are being updated using PSO algorithms. In each iteration, one of the parameters set are being updated i.e. in first iteration for example a_is are updated then in second $b_i$iteration are updated and then after updating all parameters again the first parameter update is considered and so on [11].

Evaluation Criteria

The performance of the proposed approach will be evaluated by measuring the estimation accuracy. The estimation accuracy can be defined as the difference between the actual and estimated values. The first typical fitting criterion (MSE) is defined as in Equation 11:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (11)$$

where N is the total number of data, y is actual target value, and $\hat{y}$ its estimated target value.

The experiments will be implemented many times to ensure that MSE converges to a minimum value.

The initial values for weights will randomly be assigned within the range [-1; 1]. The training accuracy is expressed in terms of the mean absolute error, standard deviation (SD) and root mean

squared error (RMSE). The absolute mean error (ME) is expressed as

$$e_i = |P_{measured} - P_{simulated}|, \quad \bar{e} = \frac{1}{N}\sum_{i=1}^{N} e_i, \tag{12}$$

where terms *measured* and *simulated* denote received signal strength that are obtained by measurement and simulated by ANFIS, while *N* is total number of samples. Standard deviation is given by

$$\sigma = \sqrt{\frac{1}{N-1}(e_i - \bar{e})^2} \tag{13}$$

The root mean squared error (RMSE) is calculated according to the expression

$$RMSE = \sqrt{\sigma^2 + \bar{e}^2} \tag{14}$$

III. RESEARCH METHODOLOGY

A. *Practical Measurement of P_R*

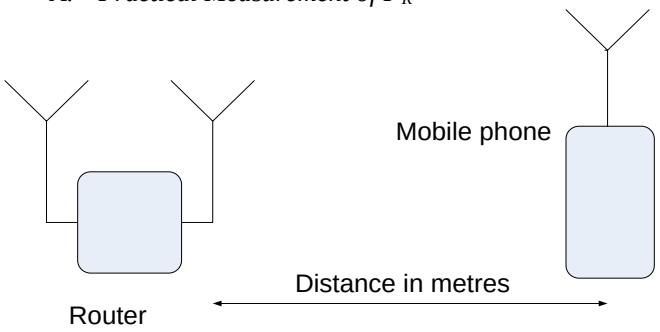


Fig. 2: Diagram of the experimental Set up

Fig. 3: Image for the experimental Set up

Fig. 4: Image for the experimental Set up in a corridor

The steps for carrying out the experiment are as follows;

i. A tape measure was used to measure a distance of 42m that was subdivided into 42 points each 1m apart.
ii. The Tecno R7 mobile device was moved metre by metre away for the D-link router and took the readings for every 1m from the router in Table 1.

B. *Data analysis*

For this study, the content analysis technique was employed to analyze the data. Matlab graphical representation techniques were used to analyze quantitative data. The full analysis on the key findings of this study is presented in the section below.

IV. FINDINGS AND DISCUSSIONS

A. *Results*

For the LOS case, the results were as shown in fig. 5 below;

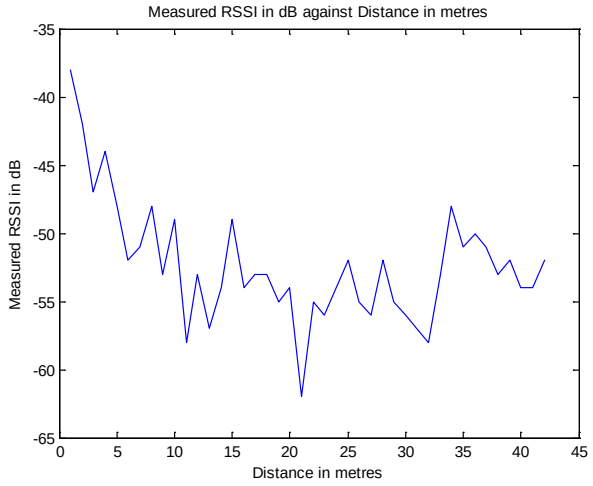


Fig. 5: LOS received signal variation with distance

Based on the measurement and Matlab analysis, the following graphs were generated for training and testing.

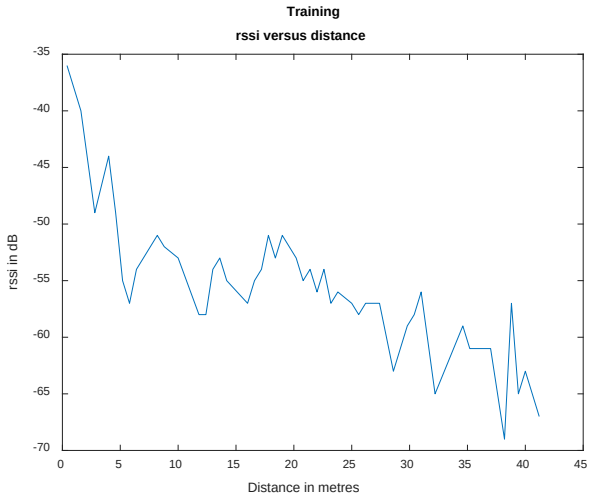


Fig. 6: Training LOS received signal variation with distance

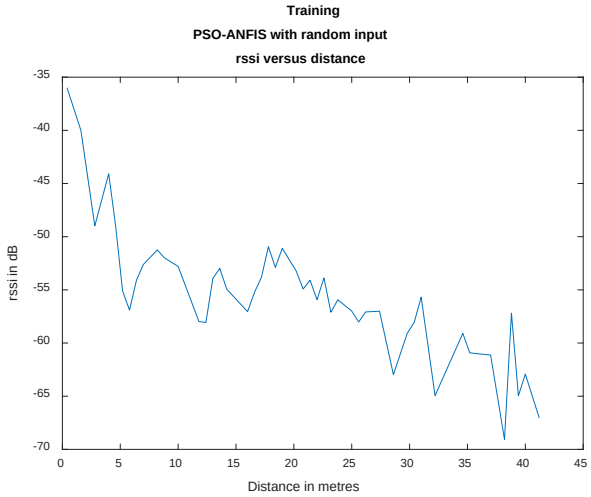


Fig. 7: Training Predicted signal variation with distance

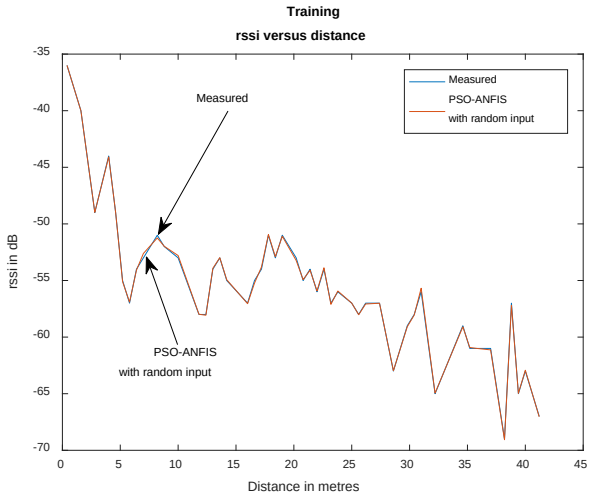


Fig. 8: Training Predicted and measured received signal variation with distance

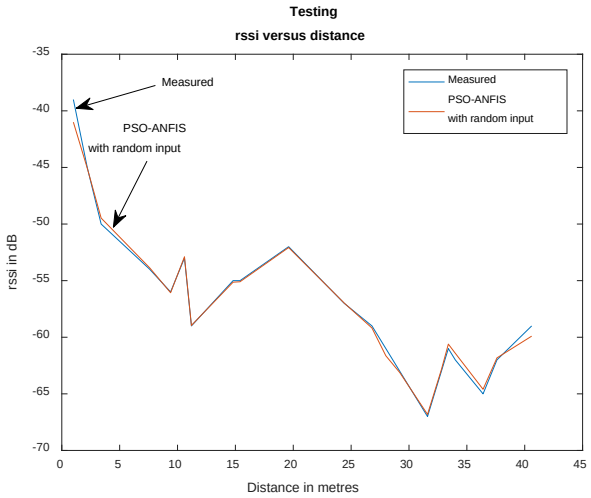


Fig. 9: Testing Predicted and measured received signal variation with distance

The graphs generated using the values obtained during the experiment and predicted are as shown above. The signal strength reduces gradually as expected due to the increase in distance between the transmitter and the receiver. For LOS propagation the time graphs show a variation in signal strength. This is due to variations in the channel conditions. The channel's transfer characteristics may vary due to movements of the transmitter, receiver or people in the indoor environment. The transmitted signal may reach the receiver through multiple reflected paths. These reflected signals may add up to strengthen each other or they may add up to cancel each other. Also, presence of objects in the path between the transmitter and the receiver also reduces the signal power arriving at the receiver. All this manifest themselves in the fluctuations in the power levels of different received signals.

This manifests in the first graph which has variations from the first to the last points.

Fig. 7 is the training predicted signal using PSO trained ANFIS prediction tool. The variation is smooth trying to follow the actual measured values. The same applies to the testing graph as shown in fig. 9. The different parameters obtained by comparing the measured and predicted values for the training and testing plots are given as;

The training, mean square error (MSE) was obtained as 0.015114, root mean squared error (RMSE) as 0.12294 and standard deviation (SD) as 0.12422 while the testing mean square error (MSE) was obtained as 0.30325, root mean squared error (RMSE) as 0.55068 and standard deviation (SD) as 0.55915.

V. CONCLUSIONS AND RECOMMENDATIONS

A. Conclusion

From experiment and calculations performed as a result thereof, it can be stated that the power of a signal transmitted in free space decreases with increase in distance from the source for both predicted and measured values.

The values obtained above indicate the closeness of predicted to the measured values indicating that the PSO trained ANFIS is very accurate in modelling wireless prediction.

B. Limitations of the study

The major limitation of the study was random behavior of the received signal.

C. Areas of further study

Future research should include the use of different training methods and compare the resulting parameters.

REFERENCES:

[1] Sohail, Ahsan, Zeeshan Ahmad, and Iftikhar Ali. "Analysis and Measurement of Wi-Fi Signals in Indoor Environment 1." (2013).

[2] Scholz, DiplIng Peter, and K. G. Kathrein-Werke. "Basic Antenna Principles for Mobile Communications." Polarization 45.45 (2000): 12.

[3] J. S. Jang, "ANFIS: Adaptive-network-based fuzzy inference system", *IEEE Trans. Syst., Man, Cybern.*, vol. 23, pp.665 -685 1993.

[4] Kennedy G & Davis B. (1999), Electronic Communication Systems, Tata McGraw Hill, 4th Ed.

[5] Gordon G. D. & Morgan W. L (1993), Principles of Communication Satellites, John Wiley & Sons.

[6] Turner, Daniel, Stefan Savage, and Alex C. Snoeren. "On the empirical performance of self-calibrating wifi location systems." Local Computer Networks (LCN), 2011 IEEE 36th Conference on. IEEE, 2011.

[7] Mongia, Puneet Kumar, and B. J. Singh. "Planning and Optimization of Wireless LANs through Field Measurements." arXivpreprint arXiv:1309.1184 (2013).

[8] S. Sumathiand and S. Paneerselvam, "Computational intelligence paradigms: theory & applications using MATLAB", CRC Press Taylor & Francis Group, 2010.

[9] Kennedy J, Eberhart R. (1995). Particle Swarm Optimization Proc of IEEE International conference on Neural Network, Perth, Australia, IEEE Service Center Piscataway NJ,1995:1942-1948.

[10] J. Kennedy and R. C. Eberhart, "Particle swarm optimization", Proc. IEEE Int. Conf. Neural Networks, pp.1942 -1948 1995.

[11] R. Hassan, B. Cohanim, and O. de Weck, "A comparison of particle swarm optimization and the genetic algorithm," presented at the AIAA/ASME/ASCE/AHS/ASC 46th Struct., Struct. Dyn. Mater. Conf., Austin, TX, Apr. 2005.

[36] V. S. Ghomsheh, M. A. Shoorehdeli and M. Teshnehlab, "Training ANFIS structure with modified PSO algorithm", Proc. *15th Mediterranean Conf. Control Automation*, 2007.

[12] J. Kennedy and R. C. Eberhart, "Particle swarm optimization", *Proc. IEEE Int. Conf. Neural Networks*, pp.1942 -1948 1995.

[13] S. Sumathiand and S. Paneerselvam, "Computational intelligence paradigms: theory & applications using MATLAB", CRC Press Taylor & Francis Group, 2010.

[14] R. Hassan, B. Cohanim, and O. de Weck, "A comparison of particle swarm optimization and the genetic algorithm," presented at the AIAA/ASME/ASCE/AHS/ASC 46th Struct., Struct. Dyn. Mater. Conf., Austin, TX, Apr. 2005.

Comparing the performance of ANFIS, PSO-ANFIS and PSO-ANFIS with random input in indoor Wi-Fi Signal propagation prediction

Omae M. O, Ndungu E. N and Kibet P. L.

***Abstract*-**With the increase in the use of mobile devices fitted with wireless local area networks (WLANS) technologies there is need for accelerated studies on these systems to improve on the quality of service (QoS) provided to the users. Different methods have been used in signal modeling including deterministic and empirical models. This study is aimed at comparing the performance of predicting Wi-Fi signal propagation along a corridor using Particle Swarm Optimization (PSO) trained Adaptive Neural Fuzzy Inference System (ANFIS), ANFIS and PSO trained ANFIS with a random input. The mean square error, root mean square and standard deviation of the predicted signal were determined and compared. The study was undertaken using a Wi-Fi router as the transmitter and a mobile phone as the receiver in the process of data collection. The measured values were then used in the modeling. It was found that the predicted values based on PSO trained ANFIS with a random input were close to actual measured values as from the undertaken analysis giving the best prediction.
Keywords; Wi-Fi, QoS, WLANS, ANFIS, PSO-ANFIS

I. INTRODUCTION

Wi-Fi networks form one of the largest market segments of wireless networks. Coverage in line of sight (LOS) environments is limited both by physical obstacles and structural barriers, while in built environments, the main obstacles are walls [1]. What is common for both is interference in the wireless spectrum. The most commonly used ISM bands for Wi-Fi networks are at 2.4 GHz and 5 GHz, and the signals at such high frequencies do not easily pass through the obstacles. To increase connectivity and extend coverage, Wi-Fi networks use limited transmission powers, typically up to 100 mW. This gives connectivity of a few tens of meters, even through walls. At the same time, LOS connectivity may reach significantly greater distances, causing far away nodes to interfere in very unusual patterns.

ANFIS is one of the most current techniques used in function approximation besides other very many applications like classification. The technique is obtained by combining the Neural Networks and Fuzzy Logic concepts which are based on numerical analysis and natural language respectively [3].

PSO originally by Doctor Kennedy and Eberhart in 1995, used to train ANFIS and other AI processes is based on the intelligence of swarms as they move in search of food [9].

This study investigated the prediction of signal coverage of Wi-Fi networks using PSO trained ANFIS.

Omae M. O, Ndungu E. N, Kibet P. L.School of Electrical, Electronic and Information Engineering, Department of Telecommunication & Information Engineering, JKUAT (phone: +2540722805012;+2540721366349, +2540724833749 fax: +2546752711; e-mail: m_oteri@yahoo.com,ndunguen@yahoo.com, kibetlp@yahoo.com).

A. Statement of the problem

WLANS are increasingly becoming a very important concept in our lives at home and work equally. Scientists have done various studies in regard to this technology and continue to do the same to ensure quality of service (QoS) is improved to the ever growing number of users. In view of this, the idea of also adding to the progressing research in this field led to this study. PSO trained ANFIS is commonly used in approximating functions because of their advantages that include high accuracy and better computational efficiency.

B. Research objectives

Main objective;

To Compare the performance of PSO trained ANFIS with a random input, PSO trained ANFIS and ANFIS.

Specific objectives

1. Measure signal strength with variation of distance along a corridor.
2. Obtain graphical comparisons for the performance of PSO trained ANFIS with a random input, PSO trained ANFIS and ANFIS.

II. LITERATURE REVIEW

A. Introduction

Wireless networking works by sending radio transmissions on specific frequencies where listening devices can receive them. The necessary radio transmitters and receivers are built into Wi-Fi enabled equipment like routers, laptops and phones. Antennas are also key components of these radio communication systems, picking up incoming signals or radiating outgoing Wi-Fi signals [4], [5]. Some Wi-Fi antennas, particularly on routers, may be mounted externally while others are embedded inside the device's hardware enclosure [2], [6].

ANFIS combines the advantages of both neural network and fuzzy logic in its operation resulting to a powerful tool in approximating functions [3].

PSO finds the optimal solution by simulating the social behavior of groups as fish schooling or bird flocking. A group can achieve the objective effectively by using the common

information of every particle (global), and the information owned by the particle itself (personal) [9].

B. Other methods used in Wi-Fi signal prediction

1. COST231 One-Slope Model

Empirical models describe the signal level loss by empirical formulas with empirical parameters optimized by measurement campaigns in various buildings to make the empirical parameters of the model as universal as possible. The COST231 One-Slope model (OSM) is the simplest approach to signal loss prediction, because it is based only on the distance between the transmitter and the receiver. This simplest prediction model does not take into account the position of obstacles, the influence of which is respected only by the power decay factor (2). Factor n and the signal loss at a distance d_0 from the transmitter L(d) in equation (1) increase for a more lossy environment, but they are constant for the whole building [15], [16], [17].

$$L_{OSM} = (d_0) + n10\left(\frac{d}{d_0}\right) \quad (1)$$

where: L_{OSM}.....Predicted signal loss (dB)
$L_0(d_0)$.....Signal loss at distance d from transmitter (dB)
n.....Power decay factor (-)
d.....Distance between antennas (m)
d_0.....Reference distance between antennas (usually 1 m) (m)

2. Dual-Slope Model

The path loss in dB is given by experimentally.

$$L_{dB} = L_{0,dB} + \begin{cases} 10n_1 log_{10} d, \ 1m < d \leq d_{bp} \\ 10n_1 log_{10} d + 10n_2 log_{10}\left(\frac{d}{d_{bp}}\right), \ d > d_{bp} \end{cases} \quad (2)$$

Basically, this model divides the distances into one line-of-sight (LOS) and one obstructed LOS region. The break point distance d_{bp} takes into account that in indoor environments the ellipsoidal Fresnel zone can be obstructed by the ceiling or the walls, anticipating the LOS region:

$$d_{dp} = \frac{4h_b h_m}{\lambda} \quad (3)$$

where h_b and h_m denote the shortest distance from the ground or wall of the access point (AP) and station (STA), respectively [25].

3. Partitioned Model

The path loss in dB is given by

$$L_{dB} = L_{0,dB} + \begin{cases} 20log_{10} d, \ 1m < d \leq 10m \\ 20 + 30log_{10}\left(\frac{d}{10}\right), \ 10m < d \leq 20m \\ 29 + 60log_{10}\left(\frac{d}{20}\right), \ 20m < d \leq 40m \\ 47 + 120log_{10}\left(\frac{d}{40}\right), \ d > 40m \end{cases} \quad (4)$$

This model uses pre-determined values for the path loss exponents and breakpoint distances, according to previous field measurement campaigns [15].

4. Average Walls Model

This model is based on the Cost-231 multi-wall except that the loss due to obstructing walls is aggregated in just one parameter L. Therefore, for a single floor environment, the path loss estimated by (5) is modified to

$$L_{dB} = 20log_{10} d + k_w L_w \quad (5)$$

where k_w denotes the number of penetrated walls. In order to determine the parameter L_w, each wall obstructing the direct path between the receiver and the transmitter antennas must have its loss measured as follows.
The loss of the first wall in dB is given by:

$$L_1 = L - L_{0,dB} - 20log_{10} d \quad (6)$$

Where $L_{0,dB}$ is the path loss obtained at 1 meter distant from the transmitter; L denotes the measured total loss from 1 meter distant after the obstructing wall. For the second wall the loss of the first wall also must be taken into account. Therefore, the loss in dB of the second obstructing wall can be estimated as

$$L_2 = L - L_{0,dB} - 20log_{10} d - L_1 \quad (7)$$

Keeping on the above methodology, the ith wall loss is given by

$$L_i = L - L_{0,dB} - 20log_{10} d - \sum_{j=1}^{i=1} L_j \quad (8)$$

where the sum spans the losses of walls obtained previously. After all wall losses of the environment had been obtained, then the wall losses average value is computed and assigned to the parameter L_w [15].

5. Multi-Wall Model

The OSM is insufficiently accurate for most applications, due to the usually inhomogeneous structure of building with long waveguiding corridors or large open spaces on one side and small complex rooms with many obstacles on the other side. For such cases, the more accurate, but still partly empirical, Multi Wall model (MWM) employing a site-specific building structure description can be used.
The Multi-Wall model takes into account wall and floor penetration loss factors in addition to the free space loss (9). The transmission loss factors of the walls or floors passed by the straight-line joining the two antennas are cumulated into the total penetration loss L_{Walls} (10) or L (11), respectively. Depending on the model, either homogenous wall or floor transmission loss factors or individual transmission loss factors can be used. The more detailed the description of the walls and floors, the better the prediction accuracy. The penetration losses are optimized as other empirical parameters from measurements, so they are not equal to the real obstacle transmission losses, but only correspond to the appropriate empirical attenuation factors of the obstacles.

$$L_{MWM} = L_1 + 20log_{10}(d) + L_{Walls} + L_{Floors} \ldots\ldots\ldots\ldots..(9)$$
$$L_{Walls} = \sum_{i=1}^{I} a_{w_i} k_{w_i} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(10)$$
$$L_{Walls} = a_f k_f \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(11)$$

L_{MWM}.....Predicted signal loss (dB)
L_1.....Free space loss at a distance of 1m from transmitter (dB)
L_{Walls}.....Contribution of walls to total signal loss (dB)
L_{Floors}.....Contribution of floors to total signal loss (dB)
a_{wi}.....Transmission loss factor of one wall of i-th kind (dB)
k_{wi}.....Number of walls of i-th kind (-)
I.....Number of wall kinds (-)
a_f.....Transmission loss factor of one floor (dB)
k_f……Number of floors (-)

Since the MWM considers the positions and specific transmission loss factor of walls, its results are more accurate than those of OSM. However, the shadowing effect of more closely adjacent walls are often overestimated, because their cumulated transmission loss factors lead to very small values of predicted signal level behind these elements. In other words the real signal may not follow a straight-line between antennas, but it can go around the walls. The computation time of the MWM is also quite short, and the sensitivity of the model to the accuracy of the description of the building is limited due to the simple consideration of only the number of obstacles passed by a straight line.

6. Artificial Neural Networks

According to [2] indoor radio propagation is a very complex and difficult radio propagation environment because the shortest direct path between transmit and receive locations is usually blocked by walls, ceilings or other objects. Signals propagate along the corridors and other open areas, depending on the structure of the building. In modeling indoor propagation, the following parameters must be considered: construction materials (reinforced concrete, brick, metal, glass, etc.), types of interiors (rooms with or without windows, hallways with or without door, etc.), locations within a building (ground floor, n^{th} floor, basement, etc.) and the location of transmitter and receiver antennas (on the same floor, on different floors, etc.). An alternative approach to the field strength prediction in indoor environment is given by prediction models based on artificial neural networks.

During last years, Artificial Neural Networks (ANN) have experienced a great development. ANN applications are already very numerous. Although there are several types of ANN's all of them share the following features: exact analytical formula impossible; required accuracy around some percent; medium quantity of data to process; environment adaptation that allows them to learn from a changing environment and parallel structure that allows them to achieve high computation speed. All these characteristics of ANN's make them suitable for predicting field strength in different environments. The prediction of field strength can be described as the transformation of an input vector containing topographical and morphographical information (e.g. path profile) to the desired output value. The unknown transformation is a scalar function of many variables (several inputs and a single output), because a huge amount of input data has to be processed. Owing to the complexity of the influences of the natural environment, the transformation function cannot be given analytically. It is known only at discrete points where measurement data are available or in cases with clearly defined propagation conditions which allow applying simple rules like free space propagation, etc.

The problem of predicting propagation loss between two points may be seen as a function of several inputs and a single output [20]. The inputs contain information about the transmitter and receiver locations, surrounding buildings, frequency, etc while the output gives the propagation loss for those inputs. From this point of view, research in propagation loss modeling consists in finding both the inputs and the function that best approximate the propagation loss. Given that ANN's are capable of function approximation, they are useful for the propagation loss modeling. The feedforward neural networks are very well suited for prediction purposes because do not allow any feedback from the output (field strength or path loss) to the input (topographical and morphographical data).

The presented studies develop a number of Multilayer Perceptron Neural Networks (MLP-NN) and Generalized Radial Basis Function Neural Networks (RBF-NN) based models trained on extended data set of propagation path loss measurements taken in an indoor environment. The performance of the neural network based models is evaluated by comparing their prediction error (μ), standard deviation (σ) and root mean square error (RMS) between their predicted values and measurements data. Also a comparison with the results obtained by applying an empirical model is done [2]. A drawback with multilayered feed-forward networks that contain numerous neurons in each layer is the required training time. Furthermore, an overly complex ANN may lead to data overfitting and, hence, generalization problems [19].

C. Adaptive Neuro-Fuzzy Inference System (ANFIS)

Adaptive Neuro-Fuzzy Inference System (ANFIS) otherwise referred to as Adaptive Network-based Fuzzy Inference System was proposed in [7]. ANFIS is a blend of Fuzzy Logic (FL) and Artificial Neural Network (ANN) that captures the strengths and offsets the limitations of both techniques for building Inference Systems (IS) with improved results and enhanced intelligence. Fuzzy logic is associated with the theory of fuzzy set, which relates to classes of objects with rough boundaries in which membership is a matter of degree. It is an extensive multivalued logical system that departs in concept and substance from the traditional multivalued logical systems. Much of fuzzy logic may be viewed as a platform for computing with words rather than numbers. The use of words for computing is closer to human intuition and exploits the tolerance for imprecision, thereby lowering the cost of the solution [8]. However, there are no known appropriate or well-established methods of defining rules and membership functions based on human knowledge and experience. Artificial Neural Networks are made up of simple processing elements operating concurrently. These elements model the biological nervous system, with the network functions predominantly determined by the connections between the elements. Neural

Networks have the ability to learn from data by adjusting the values of the connections (weights) between the elements. Merging these two artificial intelligence paradigms together offers the learning power of neural networks and the knowledge representation of fuzzy logic for making inferences from observations.

Basic ANFIS Architecture

The ANFIS architecture described here is based on type 3 fuzzy inference system (other popular types are the type 1 and type 2). In the type 3 inference system, the Takagi and Sugeno's (TKS) if-then rules are used [3]. The output of each rule is obtained by adding a constant term to the linear combination of the input variables. Final output is then computed by taking the weighted average of each rule's output. The type 3 ANFIS architecture with two inputs (x and y) and one output, z, is shown in Fig. 1.

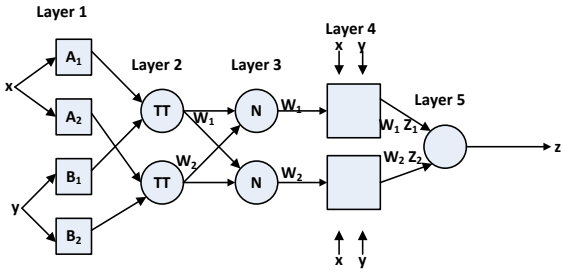


Fig. 1. Type 3 ANFIS Architecture.

$Rule\ 1: If\ x\ is\ A_1\ and\ y\ is\ B_1, then\ z_1 = p_1x + q_1y + r_1$
$Rule\ 2: If\ x\ is\ A_2\ and\ y\ is\ B_2, then\ z_2 = p_2x + q_2y + r_2$

The ANFIS structure is the functional equivalent of a supervised, feed-forward neural network with one input layer, three hidden layers and one output layer, whose functionality are as described below:

Layer 1 (Fuzzy Layer): Every node in this layer is an adaptive layer that generates the membership grades of the input vectors. Usually, a bell-shaped (Gaussian) function with maximum equal to 1 and minimum equal to 0 is used for implementing the node function:

$$O_i^1 = f(x, a, b, c) = \mu_{Ai}(x) = \frac{1}{1+|(x-c_i)/a_i|^{2b_i}}$$

$$\mu_{Ai}(x) = \exp\{-\left[\left(\frac{x-c_i}{a_i}\right)^2\right]^{b_i}\} \quad (12)$$

Where O_i^1 is the output of the i^{th}node in the first layer, $\mu_{Ai}(x)$ is the membership function of input in the linguistic variable A_i. The parameter set $\{a_i, b_i, c_i\}$ are responsible for defining the shapes of the membership functions. These parameters are called premise parameters.

Layer 2 (Product Layer): Each mode in this layer determines the firing strength of a rule by multiplying the membership functions associated with the rules. The nodes in this layer are fixed in nature. The firing strength of a particular rule (the output of a node) is given by:

$$w_i = O_i^2 = \mu_{Ai}(x).\mu_{Bi}(y), i = 1, 2 \quad (13)$$

Any other T-norm operator that performs fuzzy AND operation can be used in this layer.

Layer 3 (Normalized Layer): This layer consists of fixed nodes that are used to compute the ratio of the i[th] rule's firing strength to the total of all firing strengths:

$$\overline{w} = O_i^3 = \frac{w_i}{w_1+w_2}, i = 1, 2, \quad (14)$$

The outputs of this layer are otherwise known as normalized firing strength for convenience.

Layer 4 (Defuzzify Layer): This is an adaptive layer with node function given by:

$$\overline{w_i}z_i = O_i^4 = \overline{w_i}(p_ix + q_iy + r_i) \quad (15)$$

This layer essentially computes the contribution of each rule to the overall output. It is defuzzification layer and provides output values resulting from the inference of rules. The parameters in this layer $\{p_i,\ q_i,\ r_i\}$ are known as consequent parameters.

Layer 5 (Total Output Layer): There is only one fixed node in this layer. It computes the overall output as the summation of contribution from each rule:

$$\sum_i \overline{w_i}z_i = O_i^5 = \sum_i \frac{w_iz_i}{\sum_i w_i} \quad (16)$$

D. Particle Swarm Optimization (PSO)

PSO is a global optimization technique that was developed by Eberhart and Kennedy in 1995 [12], the underlying motivation of PSO algorithm was the social behavior observable in nature, such as flocks of birds and schools of fish in order to guide swarms of particles towards the most promising regions of the search space. PSO exhibits a good performance in finding solutions to static optimization problems where it is considered to be better than other algorithms like Genetic Algorithm [14]. It exploits a population of individuals to synchronously probe promising regions of the search space. In this context, the population is called a swarm and the individuals (i.e. the search points) are referred to as particles. Each particle in the swarm represents a candidate solution to the optimization problem. In a PSO system, each particle moves with an adaptable velocity through the search space, adjusting its position in the search space according to own experience and that of neighboring particles, then it retains a memory of the best position it ever encountered, a particle therefore makes use of the best position encountered by itself and the best position of neighbors to position itself towards the global minimum. The effect is that particles "fly" towards the global minimum, while still searching a wide area around the best solution [11]. The performance of each particle (i.e. the "closeness" of a particle to the global minimum) is measured according to a predefined

fitness function which is related to the problem being solved. For the purposes of this research, a particle represents the weight vector of NNs, including biases. The dimension of the search space is therefore the total number of weights and biases [11].

The iterative approach of PSO can be described by the following steps:

Step 1: Initialize a population size, positions and velocities of agents, and the number of weights and biases.

Step 2: The current best fitness achieved by particle *p* is set as *pbest*. The *pbest* with best value is set as *gbest* and this value is stored.

Step 3: Evaluate the desired optimization fitness function f_p for each particle as the Mean Square Error (MSE) over a given data set.

Step 4: Compare the evaluated fitness value f_p of each particle with its *pbest* value. If $f_p <$ pbest then pbest = f_p and $\text{best}_{xp} = x_p$, x_p is the current coordinates of particle p, and bestxp is the coordinates corresponding to particle *p's* best fitness so far.

Step 5: The objective function value is calculated for new positions of each particle. If a better position is achieved by an agent, *pbest* value is replaced by the current value. As in Step 1, *gbest* value is selected among *pbest* values. If the new *gbest* value is better than previous *gbest* value, the *gbest* value is replaced by the current *gbest* value and this value is stored. If $f_p <$ gbest then gbest = p, where *gbest* is the particle having the overall best fitness over all particles in the swarm.

Step 6: Change the velocity and location of the particle according to Equations 9 and 10, respectively.

Step 7: Fly each particle p according to Equation 9.

Step 8: If the maximum number of predetermined iterations (epochs) is exceeded, then stop; otherwise Loop to step 3 until convergence.

$$V_i = wV_{i-1} + acc * rand() * \left(best_{xp} - xp\right) + acc * rand() * \left(best_{xgbest} - xp\right) \quad (17)$$

Where *acc* is the acceleration constant that controls how far particles fly from one another, and *rand* returns a uniform random number between 0 and 1.

$$xp = xpp + V_i \quad (18)$$

V_i is the current velocity, V_{i-1} is the previous velocity, xp is the present location of the particle, xpp is the previous location of the particle, and *i* is the particle index. In step 5 the coordinates bestxp and bestxgbest are used to pull the particles towards the global minimum [11].

Learning by PSO

To develop an accurate process model using ANFIS, the training, and validation processes are among the important steps. In the training process, a set of input-output patterns is repeated to the ANFIS. From that, weights of all the interconnections between neurons are adjusted until the specified input yields the desired output. Through these activities, the ANFIS learns the correct input-output response behavior [11].

The way PSO will be employed for updating the ANFIS parameters is explained in this section. The ANFIS has two types of parameters which need training, the antecedent part parameters and the conclusion part parameters. The membership functions are assumed Gaussian as in equation (3.4), and their parameters are $\{a_i, b_i, c_i\}$, where a_i is the variance of membership functions and c_i is the center of membership functions (MFs). Also is b_i a trainable parameter. The parameters of conclusion part are trained and here are represented with $\{p_i,\ q_i,\ r_i\}$ [11].

Applying PSO for Training ANFIS parameters

There are 3 sets of trainable parameters in antecedent part $\{a_i, b_i, c_i\}$, each of these parameters has N genes. Where, N represents the number of MFs. The conclusion parts parameters ($\{p_i,\ q_i,\ r_i\}$) also are trained during optimization algorithm. The fitness is defined as mean square error (MSE) [11].

Parameters are initialized randomly in first step and then are being updated using PSO algorithms. In each iteration, one of the parameters set are being updated i.e. in first iteration for example a_is are updated then in second $b_i$iteration are updated and then after updating all parameters again the first parameter update is considered and so on [11], [21].

Evaluation Criteria

The performance of the proposed approach will be evaluated by measuring the estimation accuracy. The estimation accuracy can be defined as the difference between the actual and estimated values. The first typical fitting criterion (MSE) is defined as in Equation 11:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (19)$$

where N is the total number of data, y is actual target value, and $\hat{y}$ its estimated target value.

The experiments will be implemented many times to ensure that MSE converges to a minimum value.

The initial values for weights will randomly be assigned within the range [-1; 1]. The training accuracy is expressed in terms of the mean absolute error, standard deviation (SD) and root mean squared error (RMSE). The absolute mean error (ME) is expressed as

$$e_i = |P_{measured} - P_{simulated}|, \quad \bar{e} = \frac{1}{N}\sum_{i=1}^{N} e_i, \quad (20)$$

where terms *measured* and *simulated* denote received signal strength that are obtained by measurement and simulated by ANFIS, while *N* is total number of samples. Standard deviation is given by

$$\sigma = \sqrt{\frac{1}{N-1}(e_i - \bar{e})^2} \quad (21)$$

The root mean squared error (RMSE) is calculated according to the expression

$$RMSE = \sqrt{\sigma^2 + \bar{e}^2} \quad (22)$$

III. RESEARCH METHODOLOGY

A. *Practical Measurement of P_R*

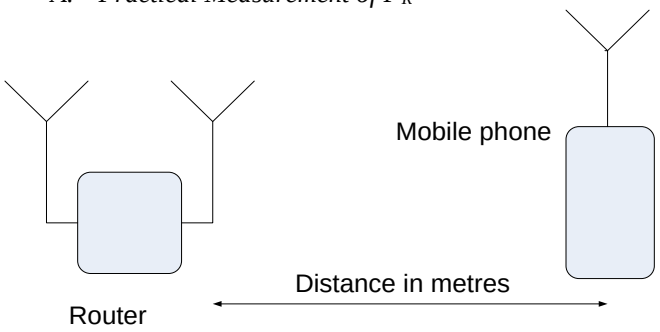


Fig. 2: Diagram of the experimental Set up

Fig. 3: Image for the experimental Set up

Fig. 4: Image for the experimental Set up in a corridor

The steps for carrying out the experiment are as follows;

i. A tape measure was used to measure a distance of 42m that was subdivided into 42 points each 1m apart.

ii. The Tecno R7 mobile device was moved metre by metre away from the D-link router and took the readings for every 1m to 42m.

B. *Data analysis*

For this study, the content analysis technique was employed to analyze the data. Matlab graphical representation techniques were used to analyze quantitative data. The full analysis on the key findings of this study is presented in the section below.

IV. FINDINGS AND DISCUSSIONS

A. *Results*

For the LOS case, the results were as shown in fig. 5 below;

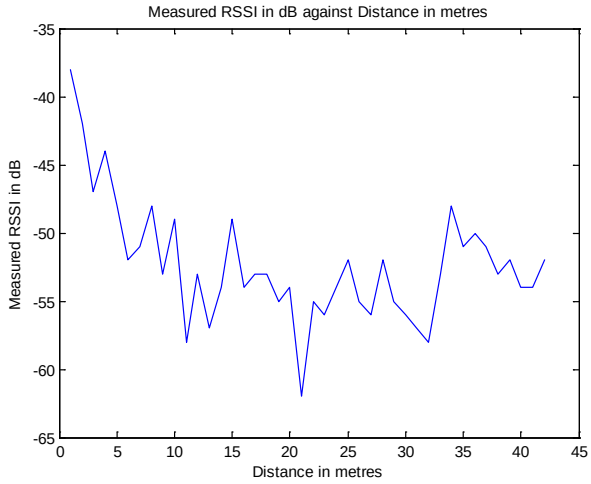


Fig. 5: LOS received signal variation with distance

Based on the measurement and Matlab analysis, the following graphs were generated for training and testing.

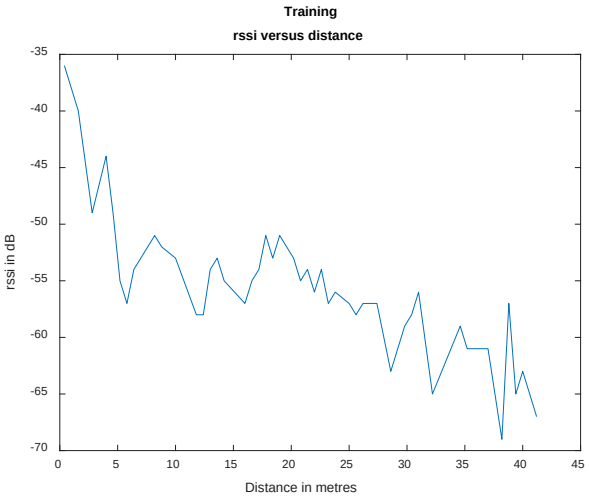


Fig. 6: Training LOS received signal variation with distance

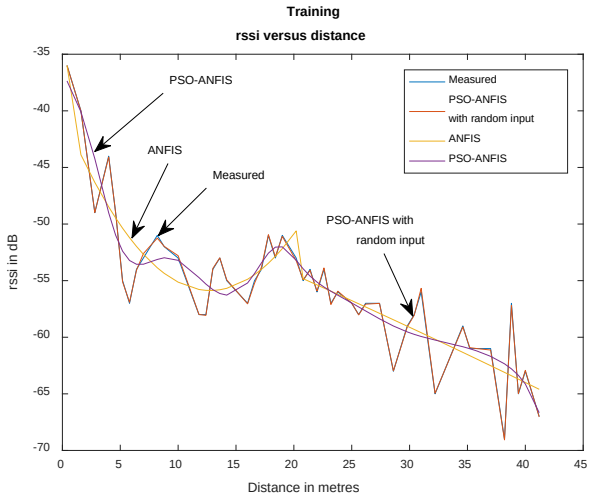


Fig. 7: Training Predicted and measured received signal variation with distance

TABLE I: TRAINING PARAMETERS

	MSE	RMSE	Standard deviation
ANFIS	6.6383	2.5765	1.6893
PSO-ANFIS	6.7002	2.3875	2.4127
PSO-ANFIS with random input	0.015114	0.12294	0.12422

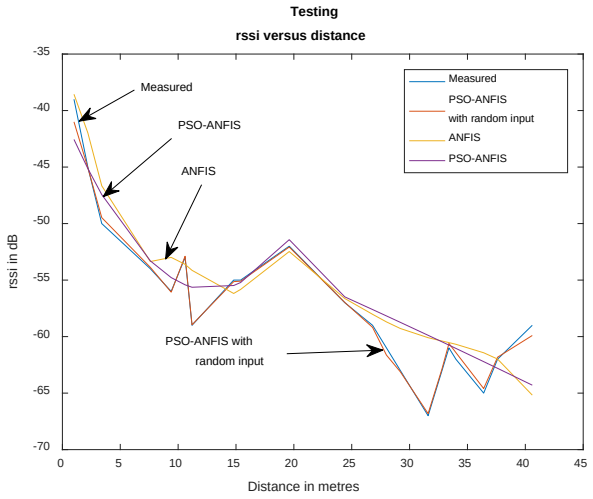


Fig. 8: Testing Predicted and measured received signal variation with distance

TABLE II: TESTING PARAMETERS

	MSE	RMSE	Standard deviation
ANFIS	7.709	2.7765	1.8682
PSO-ANFIS	8.7614	2.9583	2.6985
PSO-ANFIS with random input	0.30325	0.55068	0.55915

The graphs generated using the values obtained during the experiment and predicted are as shown above. The signal strength reduces gradually as expected due to the increase in distance between the transmitter and the receiver. For LOS propagation the time graphs show a variation in signal strength. This is due to variations in the channel conditions. The channel's transfer characteristics may vary due to movements of the transmitter, receiver or people in the indoor environment. The transmitted signal may reach the receiver through multiple reflected paths. These reflected signals may add up to strengthen each other or they may add up to cancel each other. Also, presence of objects in the path between the transmitter and the receiver also reduces the signal power arriving at the receiver. All this manifest themselves in the fluctuations in the power levels of different received signals.

This manifests in the first graph which has variations from the first to the last points.

Fig. 7 is the training predicted signal using PSO trained ANFIS with a random input, PSO trained ANFIS and ANFIS prediction tools. The variation is smooth trying to follow the actual measured values for PSO trained ANFIS with a random input tool. The same applies to the testing graph as shown in fig. 8. The different parameters obtained by comparing the measured and predicted values for the training and testing plots are given as;

The training, mean square error (MSE) was obtained as 0.015114, root mean squared error (RMSE) as 0.12294 and standard deviation (SD) as 0.12422 for PSO trained ANFIS with a random input, 6.7002, 2.3875 and 2.4127 for PSO trained ANFIS and 6.6383, 2.5765 and 1.6893 for ANFIS while the testing mean square error (MSE) was obtained as 0.30315, root mean squared error (RMSE) as 0.55068 and standard deviation (SD) as 0.55915 for PSO trained ANFIS with a random input, 8.7614, 2.9583 and 2.6985 for PSO trained ANFIS and 7.709, 2.7765 and 1.8682 for ANFIS.

V. CONCLUSIONS AND RECOMMENDATIONS

A. Conclusion

From experiment and calculations performed as a result thereof, it can be stated that the power of a signal transmitted in free space decreases with increase in distance from the source for both predicted and measured values.

The values obtained above indicate the closeness of predicted to the measured values indicating that the PSO trained ANFIS is very accurate in modelling wireless prediction.

B. Limitations of the study

The major limitation of the study was random behavior of the received signal.

C. Areas of further study

Future research should include the use of different training methods and compare the resulting parameters.

REFERENCES:

[1] Sohail, Ahsan, Zeeshan Ahmad, and Iftikhar Ali. "Analysis and Measurement of Wi-Fi Signals in Indoor Environment 1." (2013).

[2] Scholz, DiplIng Peter, and K. G. Kathrein-Werke. "Basic Antenna Principles for Mobile Communications." Polarization 45.45 (2000): 12.

[3] J. S. Jang, "ANFIS: Adaptive-network-based fuzzy inference system", *IEEE Trans. Syst., Man, Cybern.*, vol. 23, pp.665 -685 1993.

[4] Kennedy G & Davis B. (1999), Electronic Communication Systems, Tata McGraw Hill, 4th Ed.

[5] Gordon G. D. & Morgan W. L (1993), Principles of Communication Satellites, John Wiley & Sons.

[6] Turner, Daniel, Stefan Savage, and Alex C. Snoeren. "On the empirical performance of self-calibrating wifi location systems." Local Computer Networks (LCN), 2011 IEEE 36th Conference on. IEEE, 2011.

[7] Mongia, Puneet Kumar, and B. J. Singh. "Planning and Optimization of Wireless LANs through Field Measurements." arXivpreprint arXiv:1309.1184 (2013).

[8] S. Sumathiand and S. Paneerselvam, "Computational intelligence paradigms: theory & applications using MATLAB", CRC Press Taylor & Francis Group, 2010.

[9] Kennedy J, Eberhart R. (1995). Particle Swarm Optimization Proc of IEEE International conference on Neural Network, Perth, Australia, IEEE Service Center Piscataway NJ,1995:1942-1948.

[10] J. Kennedy and R. C. Eberhart, "Particle swarm optimization", Proc. IEEE Int. Conf. Neural Networks, pp.1942 -1948 1995.

[11] R. Hassan, B. Cohanim, and O. de Weck, "A comparison of particle swarm optimization and the genetic algorithm," presented at the AIAA/ASME/ASCE/AHS/ASC 46th Struct., Struct. Dyn. Mater. Conf., Austin, TX, Apr. 2005.

[12] J. Kennedy and R. C. Eberhart, "Particle swarm optimization", *Proc. IEEE Int. Conf. Neural Networks*, pp.1942 -1948 1995.

[13] S. Sumathiand and S. Paneerselvam, "Computational intelligence paradigms: theory & applications using MATLAB", CRC Press Taylor & Francis Group, 2010.

[14] R. Hassan, B. Cohanim, and O. de Weck, "A comparison of particle swarm optimization and the genetic algorithm," presented at the AIAA/ASME/ASCE/AHS/ASC 46th Struct., Struct. Dyn. Mater. Conf., Austin, TX, Apr. 2005.

[15] Andrade, C. B. and Hoeful, R. P. F. "IEEE 802.11WLANs: A Comparison on Indoor CoverageModels". In Proceedings of the 23rd Canadian Conference on Electrical and Computer Engineering, 2010.

[16] D. Xu, J. Zhang, X. Gao, P. Zhang, and Y. Wu, "Indoor Office Propagation Measurements and Path Loss Models at 5.25 GHz," IEEE Veh. Technol. Conf. (VTC), pp. 844-848, Oct. 2007.

[17] Abhayawardhana, V.S., Wassell, I.J., Crosby, D., Sellars, M.P., and Brown, M.G., "Comparison of Empirical Propagation Path Loss Models for Fixed Wireless Access Systems", VTC 2005.

[18] Caleb Phillips, Douglas Sicker and Dirk Grunwal, "A Survey of Wireless Path Loss Prediction and Coverage Mapping Methods," IEEE Communications Surveys & Tutorials, Accepted for Publication, Feb. 2012.

[19] Erik Östlin, Hans-Jürg en Zepernick a nd Hajime S uzuki, "Macrocell Path-Loss Prediction Using Artificial Neural Networks," IEEE Transactions on Vehicular Technology, Vol. 59, No. 6, July 2010, pg. 2735-2746.

[20] Popescu, I.; Nikitopoulos, D.; Constantinou, P. & Nafornita, I. (2006), "Comparison of ANN based models for path loss prediction in indoor environments," Proceedings of 64th IEEE Vehicular Technology Conference (VTC-2006 Fall), pp. 1-5, Montreal, Sep. 2006, doi:10.1109/VTCF.2006.43

[21] V. S. Ghomsheh, M. A. Shoorehdeli and M. Teshnehlab, "Training ANFIS structure with modified PSO algorithm", Proc. *15th Mediterranean Conf. Control Automation*, 2007.

Enhancement of Solar Photovoltaic (PV) Power Generation Efficiency Using Thermoelectric Generator (TEG) Modules

G. Kidegho[1], R. Kinyua[2], C. Muriithi[3] , W. Hornig[4]

Abstract—Solar photovoltaic power generation has gained wide popularity worldwide due to its renewable nature. However, high temperature conditions compromise the power generation by a substantial margin. This study was intended to investigate the effects of temperature and how these effects could be reduced using a cooling plate mounted on the backside of the PV module connected to a thermoelectric generator (TEG). Thermoelectric generators generate electricity using a temperature gradient that is created by the PV module back plate. Under varying irradiance, weather and temperature conditions the PV + TEG arrangement was used to generate power. In this set up, a 13Wp Polycrystalline 12V PV module and series connected TEG modules were monitored. The TEG modules type SP1848-271455SA, 40mm x 40mm x 3.5mm were series connected and mounted under the PV module sandwiching a graphite thermal conduction pad. An aluminium honeycomb (BPE) Alucore cooling plate was used to clamp the PV module to the TEG and then made to float in a water tank. A similar 13Wp PV module, without cooling, was monitored under the same conditions. The results showed average open circuit voltage gain of +3.5% and D.C power gain of +6% on the cooled PV+TEG module. The TEG had an average open circuit voltage of 1.63 volts with a peak of 3.6volts under high irradiance conditions. When the power generated from the TEG is taken into consideration, a much higher power gain could be achieved.

Keywords: *Thermoelectric generator, Photovoltaic, Cooling Module, Figure of merit Zt*

1.0 Introduction

High Photovoltaic (PV) module temperatures have far-reaching adverse effects on the power generation potential of solar PV modules. Considering that solar irradiation is not constant due to varying daily weather conditions, it is prudent to improve the power generation output by varying the physical conditions on the PV module. These include the temperature and the angle of tilt of the module. The main parameters of a PV module are the current and the voltage. Both are proportional to the solar irradiance. However, the voltage is inversely proportional to the module temperature. Cooling of PV modules therefore could result into more power generation by maintaining the output voltage of the PV near constant for varying temperatures.

The elevated module temperature could be exploited by using it to create a temperature gradient that is suitable for power generation in a Thermoelectric Generator (TEG) module. Thermoelectric generators are metals that are connected to form a p-n junction to generate electricity using the Seebeck phenomenon [1]. Thermoelectric generators are formed using materials that have high electric conductivity and low thermal conductivity that is known as the Figure of merit (Zt). Most of the materials commonly used have a figure of merit (Zt) ranging between 1 and 1.5. Examples include Bismuth telluride (Bi_2Te_3) Silicon germanium (SiGe) and Lead telluride (PbTe) [2].

2.0 PV and TEG Power Generation Principles

Photovoltaic modules generate electricity from the solar irradiance falling on them. The voltage/current characteristic curve of a PV module is mostly linear at low voltages up to about 15volts and becomes inverted -exponential at higher voltages as it approaches the open circuit voltage (Voc) [3]

The voltage output of a PV module at any ambient temperature can be estimated as;

$$V_{oc}(mod) = Temp\ coeff \times (T_{STC}[℃] - T(mod)[℃]) + V_{oc}(rated)[V] \quad (1)$$

Where,

Voc is the module open circuit voltage during measurement,

Temp coeff is the temperature coefficient for the polycrystalline module type ($0.12V/^0C$ [4],

T_{STC} is the temperature at standard test conditions,

T(mod) is the module temperature during the time of measurement,

Voc (rated) is the the rated factory stated open circuit voltage at STC

[1]*Gideon Kidegho, Department of Electrical & Electronic Engineering, Technical University of Mombasa, P.O Box 90420-80100, Mombasa, gguyo@tum.ac.ke*

[2]*Robert Kinyua, Department of Physics, Jomo Kenyatta University of Agriculture and Technology, P.O Box 62000-0200 Nairobi, kinyua@fsc.jkuat.ac.ke*

[3]*Christopher Muriithi, Department of Electrical & Electronic Engineering, Murang'a University of Technology, P.O Box 75-10200 Murang'a)*

[4]*Wolfgang Hornig, BPE International DR Hornig GMBH, Eckental 90542, Germany*

Table 1: The Solar PV Module Specifications at STC

PARAMETER	VALUE
P_{MPP}	13Wp
I_{MPP}	0.78A DC
V_{MPP}	18 VDC
V_{OC}	21.6VDC
I_{SC}	0.8 ADC
M_{SV}	715 VDC

On the other hand, the TEG relies on a temperature gradient between its two faces to generate electrical power. The higher the Zt of the material the higher the electrical power it shall be able to generate [5].

Zt is given as = $\left[\frac{(\alpha^2\sigma)}{k} x\ T\right]$ (2)

Where σ is the electrical conductivity, α is the Seebeck coefficient and k is the thermal conductivity of the material and T is the absolute temperature.

The voltage generated by a TEG material is given by [5]

V = S* Δt (3)

Where S is the Seebeck coefficient and Δt is Th - Tc

The TEG specification in Table 2 [6]

Table 2: TEG Specifications

TH –TC = ΔT	Voc (Volts)	I (mA)
20^0C	0.97	225
40^0C	1.8	368
60^0C	2.4	469
80^0C	3.6	558
100^0C	4.8	669

3.0 Data Collection set-ups

Two set-ups were made to collect the data for the study. One system was set up on the rooftop of the Institute of Energy and Environmental Technology (IEET-JKUAT) and the second one was setup at the solar PV workshop within the precincts of the institute.

3.1 Set-up for Irradiance and temperature measuments

The setup for collection of continuous data was done on the roof top for convenience of data collection and to avoid any chances of shading on the pyranometer used.

Figure 1: Daily irradiance measurement on the roof using a Pyranometer (Set up 1)

Table 3*: Devices installed for Temperature and Irradiance* data collection in system 1

NAME	**DESCRIPTION**
TEMPERATURE SENSOR	PT100
SOLAR PV MODULE	Polycrystalline -13Wp
PYRANOMETER	CMP 3
DATA LOGGER	COMBILOG 1022

Both systems were set up using the module type whose specifications are shown in Table 1.

3.2 Characterization of the 13Wp PV Modules

Before the setup 2 was done to measure the open circuit voltage of the PV modules, Load voltage, load current, surface temperature and the cooling water temperature, the selected modules were first checked against their technical specifications and their Voc measured under same conditions so as to characterise and compare them. This test was also done to identify any major differences on the open circuit voltage output before the temperature effects test began. The two modules were laid flat on a wooden board considering that the site of testing (Juja) is only 1.16^0 South of the equator [7]. The testing was done for 240 minutes when the weather was relatively clear as in Figure 2.

Figure 2: Characterisation of the 13Wp PV Modules

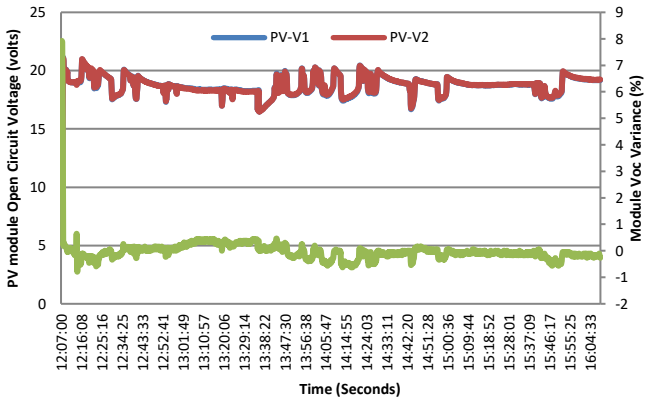


Figure 3: 13W_p PV Module Characterisation

3.3 Set-up 2

Measurement set-up 2 was made outside the solar PV workshop convenient enough to access the data logging equipment from outside. The setup was for collecting PV module open-circuit voltage, short-circuit current, load voltage, and load current. The other data collected was open circuit voltage for the thermoelectric generator, water temperature, spot irradiance and spot module temperature.

Table 4*: Major Equipment installed for Set-up 2* data collection system.

NAME	DESCRIPTION
UN-COOLED PV MODULE	Poly-13Wp
COOLED PV MODULE	Poly-13Wp
DATA STORAGE	Laptop-Lenovo
DATA LOGGER	KEYENCE-NR500
TEMPERATURE METER	AD5615
IRRADIANCE METER	TENMARS-208
MULTIMETERS	Hioki 3287 RMS

Figure 4: Set up number 2 at the solar PV workshop

The capacity of the water reservoir steel tank was 0.4392 cubic meters.

4.0 Data collection and Analysis

Data was collected from setup 1 and setup 2 simultaneously. However, the setup 2 data was collected only when solar irradiation was available during the day and in intervals of two to three hours enough for analysis while the setup1 data was continuous 24hours a day unless a deliberate interruption was necessary.

4.1 Data Collection Setup 1

The data collected in setup 1 was mostly module surface temperature and daily horizontal irradiance. This data was collected to observe the effects of direct irradiance on module surface temperature and also observe daily horizontal irradiation potential for the site (IEET-JKUAT) 1.16^0 south of the Equator.

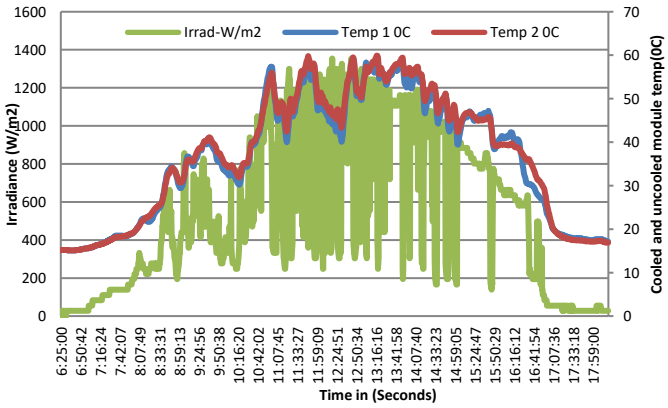


Figure 5: PV Module surface temperature and daily Irradiance

The day under observation had an irradiation energy potential of 6.32kWh/m^2 of as observed. The temperature difference between the cooled and uncooled modules was very small with an average of 1.5% due to the prevailing weather conditions in Figure 5.

4.2 Data collection for setup 2

The data collected in setup 2 was mainly open circuit voltage (Voc) for the cooled and uncooled module to observe the difference in Voc that is caused by temperature effects. The modules were connected open circuit to the KEYENCE NR500 logging device to collect this data as in Figure 4. At the same time, the TEG open circuit voltage was also monitored.

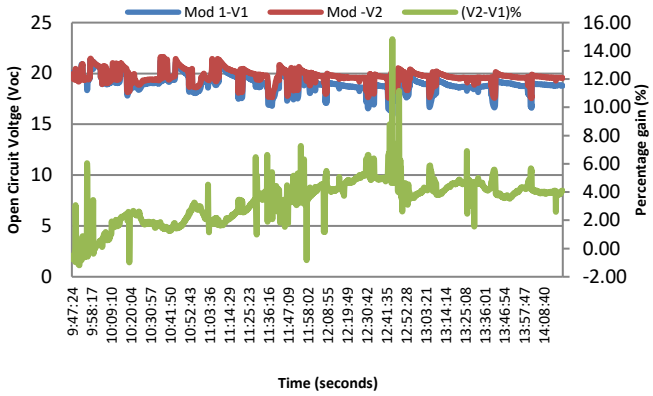


Figure 6: PV Module Open Circuit Voltage variation

The difference in Voc between the two modules was an average of 3.5% for the 13Wp module as in Figure 6. It is expected that this Voc difference would be considerable in a PV system of many series connected modules.

The other data collected in this setup was power data for the two identical modules. Both the modules were connected across a purely resistive load designed equivalent to their maximum power output load on their voltage-current characteristic curve. This was done to observe the effect of temperature on the power output of the identical modules under cooled and uncooled conditions.

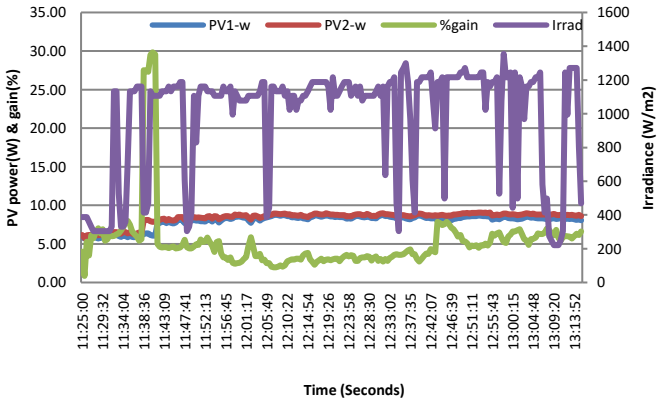


Figure 7: PV Power and Irradiance variations

In this measurement, the cooled module generated higher power output compared to the uncooled module with an average power gain of 5.4%. For the period taken during the measurement Figure 7, the cooled module harvested some 15.38Wh compared to the uncooled one of 14.64Wh with available energy of 99.61Wh

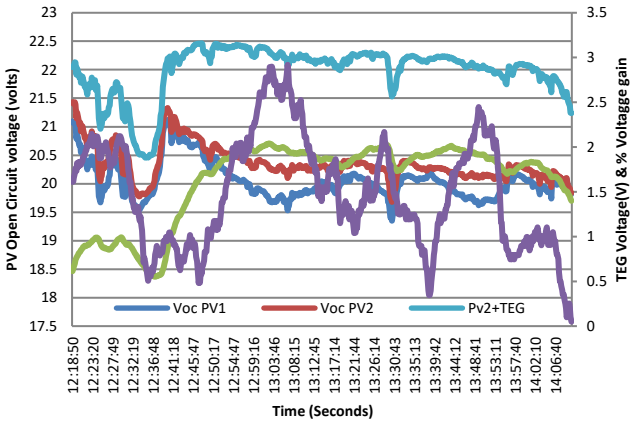


Figure 8: PV and TEG Open Circuit Voltage

In the measurement graphically presented in Figure 8, the test was done for 108 minutes when there was high irradiance. The uncooled module had an average Voc of 20.09V while the cooled module had an average Voc of 20.39. The TEG had an average Voc of 1.63V and the cooled PV module + TEG had average 22.02V. The percentage voltage gain was averaged at 1.5% for the 13Wp PV modules.

5.0 Results

The results obtained in both setups show considerable voltage and power gain with average values varying between 3.5% and 5% for the open circuit voltage and 5 and 7 % for the power gain. The energy generated for the periods observed was higher for the cooled module with an average percentage of 5.05% for the 13Wp polycrystalline module used. The TEG Voc was considerable with an average of 1.63V for the period tested. While the cooled PV module displayed a consistent higher voltage more than the uncooled PV maintaining an average percentage of +1.5%.

6.0 Conclusions

In this paper two setups for measurement were made and data was collected and analyzed to study the effects of temperature reduction on solar PV modules. The voltage output from TEG modules operating under the PV temperature gradient was observed. The open circuit voltage of the cooled PV module was found to be higher than that of the uncooled module. The energy gain averaged 5.5% for the 13Wp module. Considering that the TEG modules can be mounted under the PV modules to absorb the heat and generate electrical power, the benefits shall be considerable if the extra power generated by the TEG is quantified.

7.0 Acknowledgments:

This work is supported by the VicInAqua project. VicInAqua has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 689427.

8.0 References:

[1] Guiqiang Li; Xiao C, Jin,Y. Analysis of the Primary Constraint Conditions of an Efficient Photovoltaic-Thermoelectric Hybrid System. (2017). Energies, Vol. 10 (1), p20.

[2] Osamu, Y; Shoishi, T; Makita, K; 'Journal of applied physics', Volume 93, pp 368 -374 (2003)

[3] Roger, A.M; Jerry,V. (2010). Photovoltaic Systems Engineering' CRC press /Taylor & Francis, Boca Ranton, Vol 3, pp 49-54

[4] Partha S.P; Sudipta, M; Nasrin,A; Sharif, M. (2014). Modelling combined effects of temperature and irradiance on a solar cell parameters, 8th IEEE conference on electrical and Computer systems.

[5] Rowe, D.M. (2006). Thermoelectric Handbook: Macro to Nano, CRC press /Taylor & Francis, Boca Ranton, Vol 2, pp 21-22.

[6],Mukesh,R.Indiamart. (2018). Thermo-Electric generator. Retrieved from https://www.indiamart.com/progressiveinfoways/enquiry.html -TEG-SP1848-27145-SA-Thermo-Electric generator

[7] David B. Z. (2018). Find Latitude and Longitude. https://www.findlatitudeandlongitude.com

SNo.	Institution/Country	No. of Papers
1	Jomo Kenyatta University of Agriculture and Technology	33
2	Pan African University	5
3	Okayama University, Japan	1
4	Chiba University, Japan	1
5	Technical University of Mombasa	1
6	Tshwane University of Technology, Pretoria, South Africa	1
Total		42

JSRE is an international peer reviewed journal that publishes quarterly the original and high quality research papers with properly documented results that have major positive impact in Science and Technology, for the benefit of mankind.

The journal enjoys a global list of prestigious academic reviewers drawn from leading international universities and the editors are committed to a fair and efficient handling of manuscripts received.

We have an online manuscript management where authors can submit and track their manuscripts editorial process through our portal.

To submit your manuscript, create an account at
http://sri.jkuat.ac.ke/ojs/index.php/sri/index

Jomo Kenyatta University
of
Agriculture & Technology

Setting treds in higher learning