ISSN: 1737-9334
-PET- Vol. 53

Proceedings of Engineering & Technology
-PET-

International Conference
on Automation, Control Engineering & Computer Science

Editor:
Dr. Ahmed Rhif

International Centre for Innovation & Development –ICID–

ISSN: 1737-9334
-PET- Vol. 53

ICID

International Centre for Innovation & Development

Proceedings of Engineering & Technology
-PET-

International Conference on Automation, Control Engineering & Computer Science

Editor:

Dr. Ahmed Rhif

International Centre for Innovation & Development

–ICID–

Summary

Calculation of risk in predicting recommender system users ratings in distinct technology diffusion groups

Tranos Zuva [#1], Keneilwe Zuva[*2], Trust Tshepo Mapoka [#3]

[#] *Department of ICT*
Vaal University of Technology
Vanderbijlpark, South Africa
[1]tranosz@vut.ac.za

[*]*Department of Computer Science*
University of Botswana
Gaborone, Botswana
[2]zuvak@ub.ac.bw

#Cyber Security and Forensics
Cyber Intelligence Agency
Gaborone, Botswana
[3]ttmapoka@ciabotswana.com

***Abstract*—The number of visitors to the Internet and the rapid growth in the amount of accessible digital information have created a potential threat of information overload. Therefore timely access to items of interest on the Internet is hindered. Recommender systems are capable of mapping available content/items to user's interest, choices or detected behaviour in order to handle this issue of information overload encountered by Internet users. The system needs to have adequate information about the user to recommend relevant items to the user. This paper endeavours to group users with approximately the same risk levels of being recommended irrelevant content/items. The theory of diffusion of innovation was used in grouping users into categories. The MovieLens dataset was used in the experiment. The risk of calculating the correct rating of a user in each category was obtained then we tested whether the risk values came from different samples using F-test. The results indicated that the users in this dataset may be grouped in only three categories of the Technology Diffusion Model. We concluded that our method of knowing the users of a recommender system will improve the recommendation of relevant content/items to the users. We recommend researchers to then decide the recommender system algorithm/s to use for different categories of users being in rating prediction, ranking and/or recommendation stages.**

Keywords—recommender systems, ranking, rating prediction.

I. INTRODUCTION

The number of visitors to the Internet and the rapid growth in the amount of accessible digital information have created a potential threat of information overload. These interferes with timely access to items of interest on the Internet. Recommender systems were introduced that are capable of mapping available content to user's interest, choices or detected behaviour in order to handle this issue of information overload encountered by Internet users. Based on the user's profile, recommender system has the ability to predict whether a particular user would prefer an item or not. Recommender systems are valuable to both service providers and users(Isinkaye, Folajimi et al. 2015). They lower transaction tariffs for locating and choosing items in an online setting. They have also proven to enhance decision making process and quality (Xu and Chen 2017). In e-commerce environment, recommender systems boost earnings, for the fact that they are efficient technique of selling more products(Schafer, Konstan et al. 2001, Lee and Hosanagar 2014). In scientific libraries, recommender systems assist users by allowing them to manouver beyond catalog searches (Isinkaye, Folajimi et al. 2015). Therefore, the demand to use productive and correct recommendation approaches within a system that will provide suitable and reliable recommendations for users cannot be over-emphasized.

Recommender systems firstly solve the rating prediction problem or matrix completion problem where the system predict the rating value that a user would give for a given item at the same time completing the incomplete rating user-item matrix(Li, Singh et al. 2016, Wang, Guo et al. 2018). This problem has been researched widely and very impressive results have been obtained. Secondly the recommender system should rank the items that need to be recommended to the user (Ricci, Rokach et al. 2015). This ranking recommender system problem has been researched on and the solution tested on users in general. In most cases hybrid algorithms have been opted for in order to solve the problems encountered in using a single algorithm.

In our view different ranking algorithms should be mapped to different group of users. In recommender system we propose that ranking algorithms should take into consideration the characteristics of different groups of users of recommender systems. In this research we endeavour to group the recommender systems users in different groups using the technology diffusion theory and then calculate the risk level of each group.

The arrangement of this paper is as follows: the recommender system prediction algorithms are discussed in section 2, the ranking algorithms in section 3, proposed use of ranking algorithms in a recommender system in section 4, results in section 5 and finally conclusion.

ISSN 1737-9334

II. BACKGROUND

The purpose of recommender systems is to furnish users with personalized items, which are commonly ranked in a descending order of predicted importance. Three fundamental steps are used by these systems to make recommendations: preferences acquisition (user's input data used to acquire preferences), recommendation computation (proper methods used to compute recommendations) and recommendation presentation (the user presented with the recommendation) (Zuva, Ojo et al. 2012, Jannach and Adomavicius 2016). Existing recommendation systems can be classified into three fundamental categories based on various techniques used in recommendation computation, that is, Collaborative Filtering (CF), Content-Based Filtering (CBF) and Hybrid Filtering (HF)(Zuva, Ojo et al. 2012).

Characteristic information about items (keywords, categorifies, etc) and users (preferences, profiles, etc) are used in content based algorithms. Recommendations are made using a user's item and profile features. The assumption is that if a user was interested in an item in the past, they will once again be interested in a similar item in future. Historical interactions or explicitly asking users about their interests are used to construct user profiles. There are other systems, which utilize user personal and social data in recommending but are not considered purely content-based. Excessive specialization is one issue that has emerged that is making obvious recommendations(Jannach and Adomavicius 2016).

User-item interactions are the basis for collaborative based algorithms. These systems assume that if one item is liked by two users, then if the second user likes a second item, that very same second item could also be an interested to the first user. Hence, the objective is to use historical interactions in order to predict new ones. The similarity of users or items help to predict the next item the user might want, this makes same items to be recommended (Zuva, Ojo et al. 2012).

The main idea in using these algorithms is to predict what the users would want in their lives. The recommender system recommends an item that has a high prediction rating and if the user buys the item it is important that the user then gives feedback in order to evaluate the performance of the system. In experimenting with different algorithms there is need to have a dataset of user-ratings of items of similar nature such movies.

Cross validation model is one way of evaluating recommender systems' performance where a dataset is split into training and test data. There are many evaluation methods that are classified as cross validation models such as holdout method, K-fold cross validation and Leave-one-out validation (Ignatov, Poelmans et al. 2012).

When a recommender system algorithm has been decided, it is experimented with on a defined dataset and then implemented. A hybrid algorithm is used to make obsolete or reduce the magnitude of shortcomings and problems that might exist in other recommendation algorithm assuming that it was used individually to perform some recommendation tasks (Bela Gipp and Hentschel 2009). Users are different, therefore they should be in groups that would allow a specific algorithm to be used. Hybrid algorithms should be in place to deal with differences that exist in recommender system users(Ignatov, Poelmans et al. 2012).

The following section deals with how the recommender systems user may be partitioned into groups using theory of diffusion.

III. CLASSIFICATION OF USERS

Using theory of diffusion, the users can be put into the following categories;

A. Innovators (2.5%)

The first users to adopt an item or service are the innovators. They are willing to take risks. Their risk tolerance can lead them to adopt items or service which may sooner or later fail(Singh 2013, Aizstrautaa, Gintersa et al. 2015).

B. Early Adopters (13.5%)

The second fastest category of users who adopt an item or service are early adopters. Among the other adopter categories, these users have the highest degree of opinion leadership(Singh 2013, Aizstrautaa, Gintersa et al. 2015).

C. Early Majority (34%)

After a varying degree of time, early majority category users adopt an item or service. The innovators and early adopters time of adoption is significantly shorter than for the early majority. The adoption process of early majority tend to be slower, their social status and contact with early adopters is above average, and rarely hold positions of opinion leadership in a system (Singh 2013, Aizstrautaa, Gintersa et al. 2015).

D. Late Majority (34%)

Individuals in the late majority category will adopt an innovation after the average member of the society. These individuals approach an innovation after the majority of society has adopted it and they have a high degree of skepticism. Late majority are commonly suspicious about an innovation, their social status is below average, their financial lucidity and contact with others in late majority and early majority is very little as well as very little opinion leadership(Singh 2013, Aizstrautaa, Gintersa et al. 2015).

E. Laggards (16%)

The last adopters of an innovation are individuals in this category. They show little to no opinion leadership, this is in contrast to some of the previous categories. These individuals typically have reluctance to change-agents and gravitate towards the advanced in age. Laggards typically tend to be focused on "traditions", their social status and financial fluidity is inclined to be lowest, they turn to be the oldest as compared to all the other adopters and have contact with only family and close friends and they have very little to no opinion leadership as well(Singh 2013, Aizstrautaa, Gintersa et al. 2015).

The following diagram summarizes the categories of recommender systems users.

ISSN 1737-9334

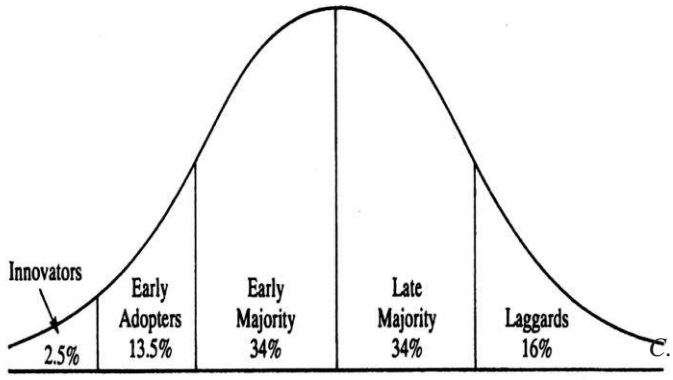


Fig. 1. Categorization of recommender systems users according to diffusion of innovation model.

The following section deals with the methodology used to partition a given recommender system dataset of user-item matrix into technology diffusion categories

IV. Methodology

In our endeavour to calculate the risk of calculating the rating prediction of users to an item/service we partitioned the recommender system users into technology diffusion categories. Dataset used is from [MovieLens] (http://movielens.org). It contains 20000263 ratings and 465564 tag applications across 27278 movies. "The data set may be used for any research purposes under the following conditions.

- The user may not state or imply any endorsement from the University of Minnesota or the GroupLens Research Group.
- The user must acknowledge the use of the data set in publications resulting from the use of the data set
- The user may not redistribute the data without separate permission.
- The user may not use this information for any commercial or revenue-bearing purposes without first obtaining permission from a faculty member of the GroupLens Research Project at the University of Minnesota".

The data set has the following heading shown in the table below:

userId	movieId	rating	timestamp

The method used to classify the recommender system users is through the timestamp column and to calculate the risk of calculating the predicated user ratings of an item/service. The following algorithm was used:

A. Sort the dataset using timestamp

B. Calculate the average time that it takes for most of the users to watch a movie.

$$\bar{X} = \frac{1}{n} * \sum_{i=1}^{n} x_i$$

where:

$\bar{X}$ = average (or arithmetic mean)

n = the number of ratings in each category

x_i = the value of each user rating in the category

C. Calculate standard deviation it takes for most of the users to watch a movie.

$$sd = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \bar{x} \right)^2}$$

where:

n = the number of ratings in each category

$\bar{x}$ = average (or arithmetic mean)

x_i =the value of each user rating in the category

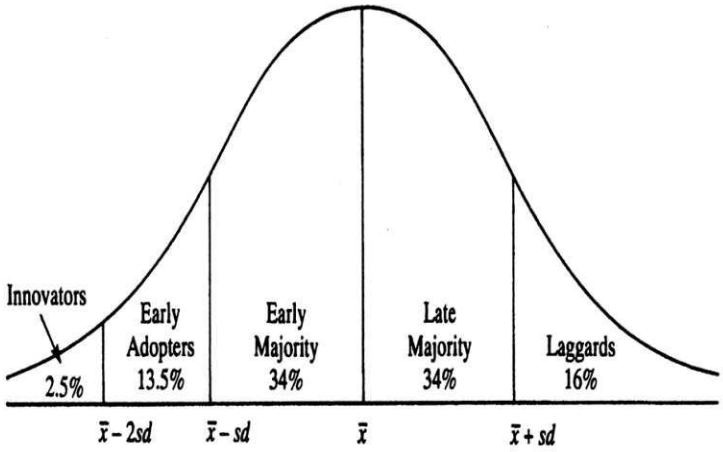


Fig. 2. Technology diffusion categories.

D. Divide the list of users using the average time calculated in B and standard deviation in C into diffusion categories as in Fig. 2.

In Fig. 2 the horizontal axis constitutes the average time which is $\bar{x}$ and sd is the standard deviation. A sample of the movies was taken using purposive sampling method. The movies that were watched by almost all the users were used in this experiment

E. Calculate the variance of their ratings in each category of diffusion

V. Result

The risk values for each category of technology diffusion are shown below:

ISSN 1737-9334

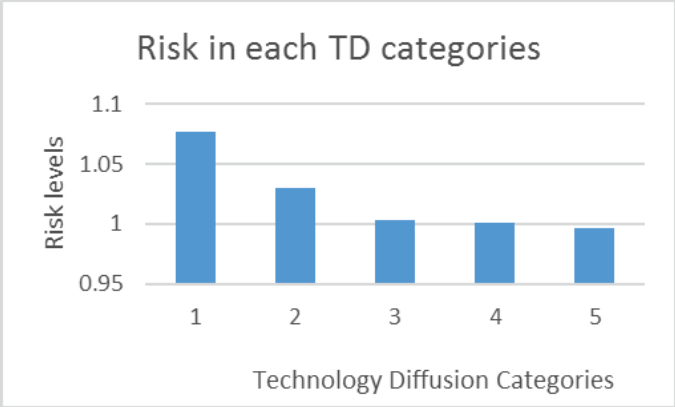


Fig. 3. Results of in each category of technology diffusion categories.

Fig. 3 has the vertical axis indicating the risk values calculated using the standard deviation and the horizontal axis indicating the technology diffusion categories. Bar 1 indicates the risk in calculating rating predication of innovators that is those users that are venturesome, educated and/or that use multiple information sources. Bar 2 indicates risk for the early adopters who are the social leaders, popular and/or educated users. Bar 3 indicates early majority who are thoughtful and/or have many informal social contacts. Bar 4 the late majority those that are sceptical, traditional and/or of lower socio-economic status while Bar 5 indicates the laggards whose neighbours and friends are main information sources and/or at the same time be afraid of debt.

In testing whether the variances of the samples of the users' ratings have a significant difference we used the F-Test. The hypothesis is that the variances are equal. In the event that if F > F Critical one-tail, we reject the null hypothesis.

TABLE I. F-TEST TWO-SAMPLE FOR VARIANCES

	Bar 1	*Bar 2*
Mean	3.699164345	3.592132505
Variance	1.076842875	1.029829319
Observations	359	1932
df	358	1931
F	1.045651795	
P(F<=f) one-tail	0.284392571	
F Critical one-tail	1.139313305*	

This is the case, F < F critical (1.045651795 < 1.139313305*). Therefore, we accept the null hypothesis. The variances of the two populations are equal.

TABLE II. F-TEST TWO-SAMPLE FOR VARIANCES

	Bar 3	*Bar 4*
Mean	*3.521783806*	*3.415330867*
Variance	*1.002711391*	*1.0013219246*
Observations	*4866*	*4866*
df	*4865*	*4865*
F	*0.885636874*	
P(F<=f) one-tail	*1.14804E-05*	
F Critical one-tail	*0.953926004**	

This is the case, F < F critical (0.885636874 < 0.953926004*). Therefore, we accept the null hypothesis. The variances of the two populations are equal.

TABLE III. F-TEST TWO-SAMPLE FOR VARIANCES

	Bar 2	*Bar 3*
Mean	3.521784	3.592132505
Variance	1.002711	1.029829319
Observations	4866	1932
df	4865	1931
F	0.973668	
P(F<=f) one-tail	0.239374	
F Critical one-tail	0.9398*	

This is the case, F > F critical (0.973668> 0.9398*. Therefore, we reject the null hypothesis. The variances of the two populations are unequal.

TABLE IV. F-TEST TWO-SAMPLE FOR VARIANCES

	Bar 4	*Bar 5*
Mean	3.415331	3.531209
Variance	1.00132192	0.996515
Observations	4866	2291
df	4865	2290
F	1.136152	
P(F<=f) one-tail	0.000209	
F Critical one-tail	1.061118*	

This is the case, F > F critical (1.136152 > 1.061118*. Therefore, we reject the null hypothesis. The variances of the two populations are unequal.

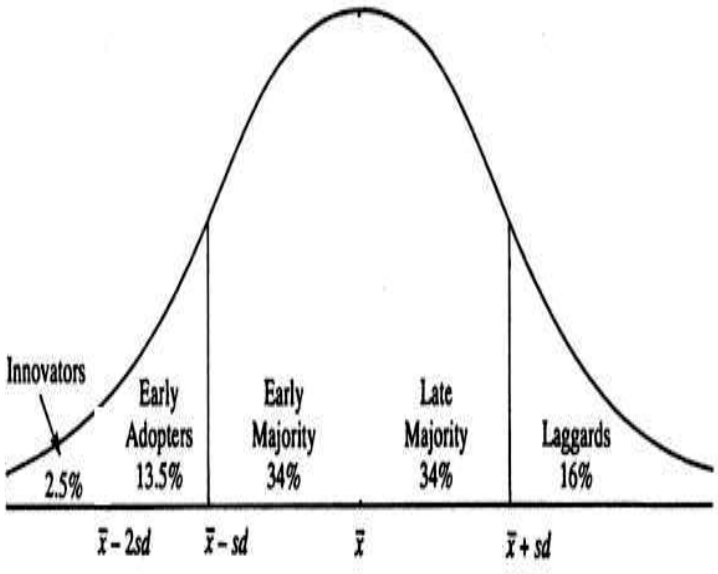


Fig. 4. According to the results recommender systems users may be categorized into these three categories

ISSN 1737-9334

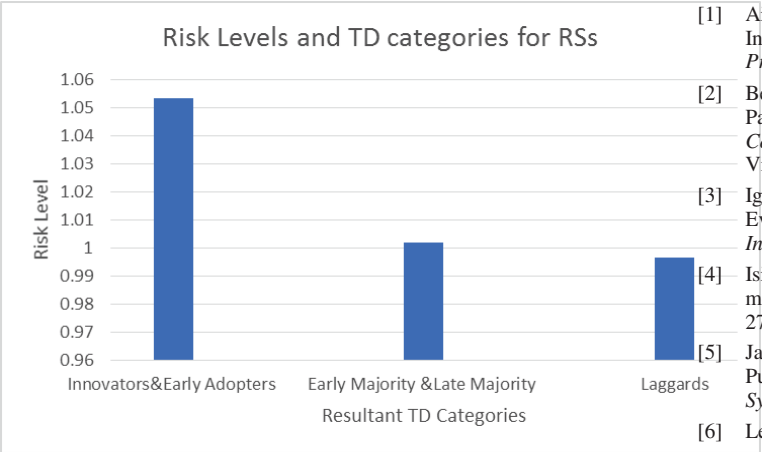


Fig. 5. Risk level in calculating prediction user ratings in technology diffusion categories for recommender systems.

CONCLUSION

In order to personalize the recommender system list so that high accuracy can be achieved on every user different ranking algorithm based and/or different recommender system prediction algorithms should be used in different technology diffusion categories. The results of the study indicate that innovators and early adopters their ratings are so wide apart or vary so much that there is need for suitable algorithm to be able to predict with higher degree of accuracy. There is a moderate variation of the ratings in early and late majority. The laggards can be predicted with some higher degree of accuracy.

REFERENCES

[1] Aizstrautaa, D., et al. (2015). "Applying Theory of Diffusion of Innovations to Evaluate Technology Acceptance and Sustainability." *Procedia Computer Science* 43(2015): 69 – 77.

[2] Bela Gipp, J. B. and C. Hentschel (2009). Scienstein: A Research Paper Recommender System. *Proceedings of the International Conference on Emerging Trends in Computing* (ICETiC'09), Virudhunagar , India.

[3] Ignatov, D. I., et al. (2012). A New Cross-Validation Technique to Evaluate Quality of Recommender Systems. *Perception and Machine Intelligence- First Indo-Japan Conference*, Kolkata, India.

[4] Isinkaye, F. O., et al. (2015). "Recommendation systems: Principles, methods and evaluation." *Egyptian Informatics Journal* 16(3): 261-273.

[5] Jannach, D. and G. Adomavicius (2016). Recommendations with a Purpose. *Proceedings of the 10th ACM Conference on Recommender Systems,* Boston, Massachusetts, USA, ACM.

[6] Lee, D. and K. Hosanagar (2014). Impact of Recommender Systems on Sales Volume and Diversity. *Thirty Fifth International Conference on Information Systems*, Auckland.

[7] Li, B., et al. (2016). Data Poisoning Attacks on Factorization-Based Collaborative Filtering. *29th Conference on Neural Information Processing Systems* (NIPS 2016), Barcelona, Spain.

[8] Ricci, F., et al. (2015). Recommender Systems Handbook. New York, Springer Science+Business Media.

[9] Schafer, J. B., et al. (2001). "E-Commerce Recommendation Applications." *Data Mining and Knowledge Discovery* 5(1): 115-153.

[10] Singh, V. (2013). Innovation diffusion categories and innovation-related needs. *Proceedings of the International Conference on Engineering Design.*

[11] Wang, Z., et al. (2018). Matrix Completionwith Preference Rankingfor Top-N Recommendation. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*(IJCAI-18), Stockholm, Sweden.

[12] Xu, B. and J. Chen (2017). "Consumer Purchase Decision- Making Process Based on the Traditional Clothing Shopping Form." *J Fashion Technol Textile Eng* 5(3): 1-12.

[13] Zuva, T., et al. (2012). "A Survey of Recommender Systems Techniques, Challenges and Evaluation Metrics." *International Journal of Emerging Technology and Advanced Engineering* 2(11): 382-386.

ISSN 1737-9334

IT Risk Modeling for Effective Information Security Management

Meltem Mutluturk #1, Bertan Badur #1 Bilgin Metin#1,

#1 *Management Information Systems Dept. Bogazici University*
Bebek, Istanbul, Turkey

1meltem.mutluturk@boun.edu.tr

1bertan.badur@boun.edu.tr

1bilgin.metin@boun.edu.tr

Abstract— **Information security is a long process and requires effective collaboration in organizations. Information Security Management Systems (ISMS) must empower collaboration between people or work groups considering risk management for information systems. There are various studies presented in order to help organizations during the ISMS process or in a general manner for information security issues. Most of them are concentrated on a specific issue of information security. Although using an approach which concentrates on a specific area can be useful, using a holistic approach helps in all aspects of business process beside information security. Organizations can use the presented approach for task allocation, risk management, policy management, monitoring and reporting. This system focuses on the contribution of people who take place in the audit process and empowers collaboration between people with the help of task module.**

Keywords— **Risk modeling, risk management, information security, decision support system**

I. Introduction

From the management's point of view, risk is the effect of uncertainty on organizational objectives which is caused by internal and external factors. Risk management is coordinated activities to direct and control and organization with regard to risk. Risk is generally calculated as the product of probability and impact. However, there are a variety of different formulas that are applied according to different situations.

Although risk assessment is a popular subject in both academical and business environment, there is very little resources on automation or use of information systems in the risk assessment process. [1] provides a framework for automated risk assessment tools, while [2] develops a functional model for risk assessment support systems.

One of the impacts that automated information system tools have on the risk assessment process is that it reduces the work effort put up by risk analysts and auditors up to 95% [1]. With the help of a software designed for risk assessment, quantifying the potential impact and the likelihood of occurrence of the risk elements would also improve. Other advantages include better representation and reporting of assessment results, ability to update risk elements after action is taken to mitigate those risk elements and easier data collection from participants of the assessment process [3-5].

This study aims to describe the need for risk modeling and use of information systems in the risk assessment process.

II. Literature Review

There are several studies presented for risk management in the literature [1-9]. From a risk automation point of view, [6] exploit the potential benefits that big data offers for the risk assessment process. Also, some studies present standards for risk management [7-9].

The literature also presents many different frameworks and standards to information security management. Many of these frameworks are based on frameworks and standards such as ISO 27001 and 27002, The National Institute of Standards and Technology (NIST) Cybersecurity Framework and Control Objectives for Information and Related Technologies (COBIT).

III. Presented Method

The presented IT risk model "Information Technology Audit Process Management Modeling" (ITAuditPM) could be very useful to support information security enhancement attempts of companies. The project aims to manage and trace the compatibility process in an effective collaboration in companies which takes ISO 27001 as s reference standard [10]. Therefore, the model focuses on the contribution of people who take place in a step of the audit process.

Information security is mainly managed by one department. Other departments support and contribute this process in different extent. For example, IT department and marketing department normally does not have equal impact on the information security process in an organization. We think a company which has an internal audit department which executes information security objectives with the help of other departments may have a similar system interaction. Context diagram in Fig. 1 shows all external entities and inputs and outputs.

As shown in Fig. 1, there are five modules in the presented model. The presented IT risk model allocates audit tasks to the people in the risk audit team and stores the findings in a database, monitors the ISO 27001 compatibility that is global standard for information security management system. Also, it manages assets and related threats in order to conduct a risk

ISSN 1737-9334

assessment. Furthermore, it stores and presents security policies and generates audit process reports. In addition to these features, it should be multiuser and only registered users should be able to use it. In the system there should be a user role structure and different users can view different aspects of the system with task allocation. The system should assign tasks to the users; these tasks consist of an audit control standard that is ISO 27001 in our case.

Furthermore, the entity-relationship diagram of ITAuditPM is shown in Fig. 2. All task findings can be traced with the help of the system and standard compliance values are determined in terms of finding value. Task owners store their findings to the system, and they can also upload related files to the system. After evaluating findings, the task owner enters a compliance value with regards to the control of the task. These values finally determine the compliance value of the organization.

On the other hand, in the presented system, task allocation works with approvals. For example, if a task is assigned to a user, the system notifies the user about this task. The user accepts and begins to work or rejects the task. All actions of the task owner are submitted to other users who create this task. After working on the given task, it can be ended. The task owner sends the task for verification of findings. User input data can be validated in a logical and reliable perspective in order to provide system integrity.

Risk assessment is also critical for many organizations today. To protect assets, their value along with the risk associated with them needs to be known [11]. ITAuditPM stores assets and possible threats for them. Assets can be registered by the responsible person and most importantly, an asset value is determined by the system manager. The system manager is generally the IT audit manager. Asset values must be evaluated sensitively with regard to not only departmental but also companywide factors. When determining an asset value, user enters availability, confidentiality and integrity values.

Also, user determines threat values and its measures. Threats have two measures to evaluate them such as Impact and Likelihood.

In order to make a correct risk assessment, the scales of the measurement elements are very crucial. Scales are defined in the system and the user selects a level for every measure. All of the assessments are kept in the system in order to view after. The system gives resolving possibility after editing threat or asset values. The system has multiple user types. These user types have different interactions with the system.

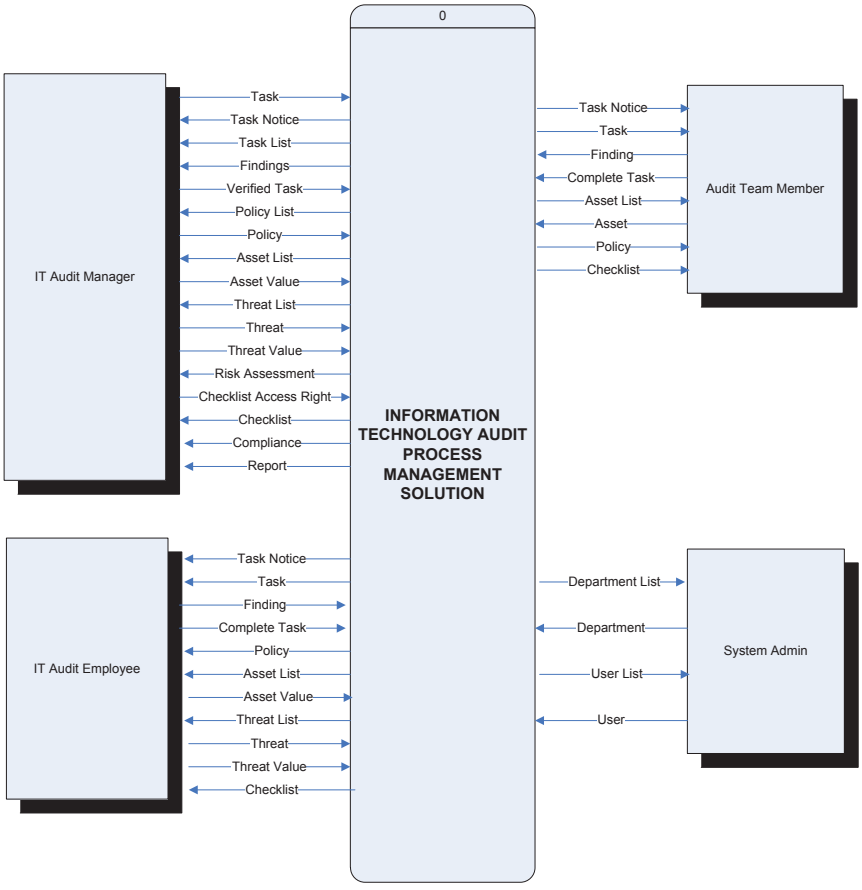


Figure 1 Context Diagram

ISSN 1737-9334

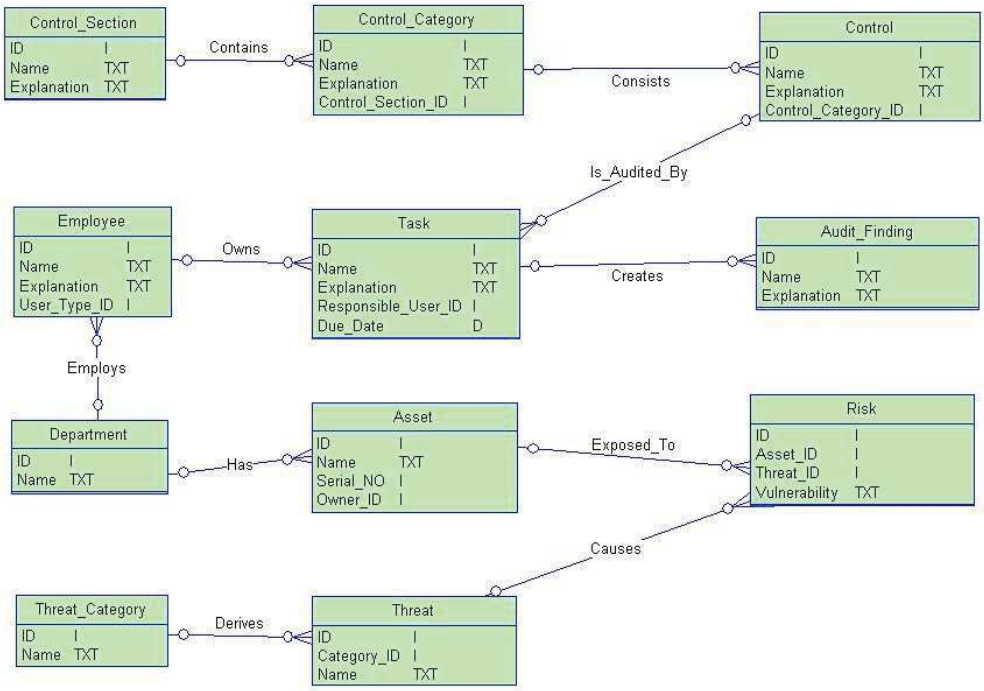


Figure 2 Entity-Relationship Diagram

The system must store and present critical data whenever needed so it must run as a reference source for information security issues. The system assigns and stores tasks that are related to a security objective. Also, system stores assets and threats in order to make risk assessments. Storing risk assessment reports is important because organizations may need to compare it with changing risk values. Another important data is security policies. Security policies help other departments act in harmony. For example, a new employment policy is beneficial for human resource departments. Human resource department can evaluate and employ a person with regard to the security policy which has been prepared in advance. Also, data libraries can support people in working on a security objective. For example, a reference standard library such as ISO 27001 helps employees during their tasks. ITAuditPM manages the data mentioned above.

IV. Conclusion

The presented Information Technology Audit Process Management Solution is suitable to support information security enhancement attempts of companies. The study aims to manage and trace compatibility processes in an effective collaboration in companies which takes ISO 27001 as a reference standard. Therefore, presented IT risk modeling focuses on the contribution of people who take place in a step of the audit process.

Acknowledgement

Authors would like to thank Gökdal Alıcı for his contribution to this study as part of his bachelor graduation project.

References

[1] Will Ozier, A Framework for Automated Risk Assessment Tools, EDPACS: The EDP Audit, Control, and Security Newsletter 2002.
[2] Artis Teilans, Andrejs Romanovs, Yuri Merkuryev, Arnis Kleins, Pjotrs Dorogovs,Ojars Krasts, Functional Modelling of IT Risk Asssessment Support System, Economics and Management, 2011.
[3] Deborah Everest-Hill, David Young, Automating Risk Assessment, Internal Auditor 1999.
[4] Satyanandan B. Atyam Effectiveness of Security Control Risk Assessments for Enterprises: Assess on the Business Perspective of Security Risks, Information Security Journal: A Global Perspective, 2010
[5] Arben Mullai, Risk Management System – Risk Assessment Frameworks and Techniques, DaGoB publication series, 2006
[6] Ankur Pariyani, Ulku G. Oktem, Deborah L. Grubbe, Process risk assessment uses big data, Control Engineering, 2013
[7] ISO 31010:2009, Risk Management – Risk Assessment Techniques
[8] ISO/IEC 27005:2011, Information Security Risk Management
[9] BS 7799-3:2017 Information security management systems. Guidelines for information security risk management
[10] IISO/IEC 27001:2013 information technology. Security techniques. Information security management systems. Requirements
[11] In, H. P., Kim, Y. G., Lee, T., Moon, C. J., Jung, Y., & Kim, I. (2004, October). A security risk analysis model for information systems. In Asian Simulation Conference (pp. 505-513). Springer, Berlin, Heidelberg.

ISSN 1737-9334

Vision based EKF SLAM in a light intensity Static Environment

Jabulani K. Makhubela[#1], Tranos Zuva[*2], Olusanya Yinka Agunbiade[#,*3]

[#]*Department of Computer Science, Tshwane University of Technology*
Pretoria, South Africa
[1]makhubelaJK@tut.ac.za
[3]sanyaYinka@gmail.com

[*]*Department of Information and Communication Technology, Vaal University of Technology*
Vanderbijlpark, South Africa
[2]tranosz@vut.ac.za

***Abstract*— Simultaneous localization and mapping (SLAM) is when an autonomous mobile robot is able create a model of an unknown environment and estimates its pose at the same time. Visual Simultaneous Localization and Mapping (VSLAM) arise when an autonomous mobile robot is embedded with a vision sensor such as monocular, stereo vision, omnidirectional or Red Green Blue Depth (RGBD) camera to localize and map an unknown environment. VSLAM has been a common technique for autonomous navigation and numerous studies by various researchers have shown tremendous improvement towards the technique. However there is need to improve on this VSLAM based on the problems encountered by other researchers as it is revealed in literature. The problem of environmental noise such as light intensity has been an issue that make VSLAM system to be ineffective. The purpose of this study is to address this challenge by introducing image filtering algorithm, together with the Extended Kalman Filter (EKF) algorithm for localization and mapping and A* algorithm for navigation into the VSLAM technique to improve the robustness of the system in a static environment. The methodology used to perform experiment in research is simulation. Experimental results show a root mean square error (RMSE) of 0.13m, which is minimal when compared with other SLAM systems from literature. The inclusion of an Image Filtering Algorithm has enabled the VSLAM system to navigate in a noisy environment.**

Keywords—Navigation, Sensor, Vision, light intensity, mobile robot, Simultaneous Localization and mapping (SLAM)

I. INTRODUCTION

The capability of an autonomous mobile robot to create a model of an unknown surrounding and estimate its position at the same time is described SLAM [1]. SLAM technique has been common challenge for autonomous robots to navigate and accomplish mobile operational activities [2]. According to [3] technique has been a prosperous for solving autonomous robot shortcomings for research community**;** it has attracted many research studies, and main reasons for robotics accomplishment [3].

The key reason for SLAM technique success it is better than others in removing the artificial infrastructures of the location [3]. For an autonomous robot to navigate the environment, majority of systems use wide range of sensors such as laser scanner, sonar, acoustic etc. which can be very large and expensive [3]**.** [4] Endorses the operation of camera into the SLAM technique, which can deliver extra information on the location than other sensors**.** [3] State that the benefits of a camera are compact, precise, low-priced, non-invasive and pervasive**.** Camera as vision sensors are favored because people and animals seems to be navigating effectively in a complicated locations using a vision as prime sensor [4].

Various researchers [5] have presented leading methods in SLAM technique such as extended Kalman filter (EKF), FastSLAM, GraphSLAM, and Rao Blackwellized Particle filter (RBPF) SLAM. Numerous researcher have concentrated on making the vision SLAM technique to operate on more, challenging environment [6]. However the problem of environmental noise like light intensity in static environment has been a matter that causes error on the VSLAM technique [7].

In the attempt to solve this challenge in our study, we have introduce image filtering algorithm into the VSLAM technique to increase the effectiveness of the system. The outcome of this study has improved VSLAM to enable an autonomous mobile robot application to navigate the static environment where environmental noise exists, without causing an error. A reminder of this paper is organized as follows: - section II discusses the methodology of proposed VSLAM technique. Section III discuss experiments & results and finally conclusion is drawn in section IV.

II. METHODOLOGY OF A PROPOSED VSLAM TECHNIQUE

For an autonomous robot to localize its position and successfully build a map in unfamiliar location, it needs to follow a method called SLAM [7]. VSLAM arise when a mobile robot uses a vision sensor as their eye(s) to map and navigate the environment [2]. In our effort to address the challenge of environment noise such as light intensity in VSLAM, we have chosen simulation as our methodology in this study.

"Simulation is defined as a method for using computer software to model the operation of real world processes, systems, or events" [8]. The proposed VSLAM technique in a noisy environment is simulated by using Matlab. Matlab stands for Matrix Laboratory, a fourth generation programming language commonly utilized for technical calculation, programming and conceptualization of data [9].

The VSLAM proposed technique comprised of five phases: Image Acquisition Phase, Pre-processing Phase, Light Filtering Algorithm, Localization and Mapping Phase and Navigation Phase. Figure 1 shows the graphical structure of the proposed VSLAM technique. The phases below will be discussed in detail later in the section. This

ISSN 1737-9334

section also discusses the dataset that is used to simulate the proposed VSLAM technique and evaluation metrics.

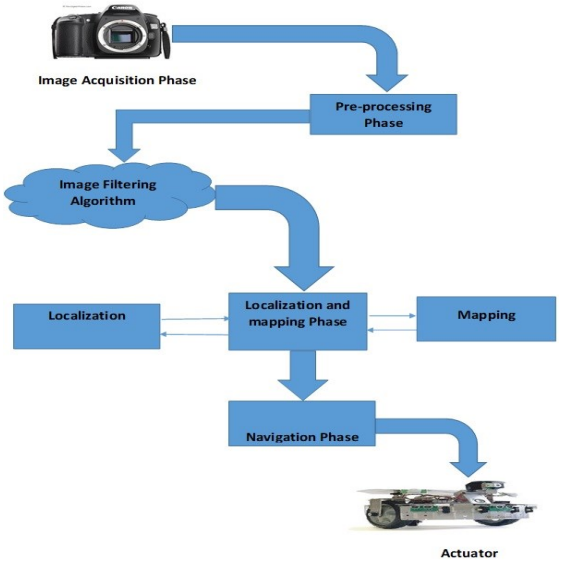


Fig 1: Proposed Framework of VSLAM technique [3]

A. Sensors (acquisition stage)

Before image processing can start, a picture must be received by a sensor and transformed into a controllable object; this method is called image acquisition [10]. According to [10] the image acquisition procedure involves three phases: energy reflected from the entity of concern, an optical system which emphasizes the energy and a sensor which calculates the quantity of energy. The energy of interest is normally electromagnetic waves [10] and the light reflected from the object has to be taken by the sensor. If a material sensitive to the reflected light is located near to the entity, a picture of the entity will be taken [10]. The light reflected from the object of concern is focused by some optics and now must be logged by the sensor [10]. The image will then be sent to pre-processing stage for further processing.

B. Pre-process Algorithm

The second phase in our method is Image pre-processing, since the major part of the genuine information from camera sensor are noisy, incomplete and inadequate, so pre-processing from acquisition stage becomes essential [11]. Image pre-processing was one of the preparatory stages which were very required to guarantee the high exactness of the subsequent stage [11]. [12] defined Image Pre-processing as a procedure to enhance crude pictures caught from cameras/sensors situated on satellites, space tests and air ships or pictures caught in standard day by day life for various frameworks. The reason for pre-pre-processing was an upgrade of the picture data that overwhelms unfavorable misrepresentations or enhances some picture features pertinent for extra processing with and analysis work [13]. The photos taken from the vision sensors experience pre-process stage to be prepared, analyzed for an autonomous robot to extract the landmarks on the picture. In our work to improve VSLAM, at the pre-processing stage, we had introduced a noise filtering algorithm to remove the noise on pictures which were harmed by higher light intensity in the Acquisition stage. The next section introduces the filtering algorithm to remove the environmental noise such as light intensity had been removed from acquired image.

C. Image Filtering Algorithm

An Image filtering algorithm is vital to our technique as it is utilized to lessen environmental noise, clean and improve the pictures that are as of now influenced by environmental noise, for example, light intensity, shadow, rain and so forth. Environmental noise, for example, light variety are common noise most particularly in office territories [7] and can cause the VSLAM method to fail, because as mentioned early in the study the environmental noise can harm picture, degenerate the Red Green Blue (RGB) shading value and bring decreased vision [14]. As mentioned early, in our pre-processing stage we had introduced a noise filtering algorithm, but in this research our attention is only focused on light intensity because of its common occurrence on a daily basis. Filtering algorithms are vital method in image processing, they are used to decrease noises in image. Environmental noise has the competency to destroy image, corrupt the RGB color value and bring poor vision that makes the image content interpretation difficult to analyses [14]. According to [14] they are several types of Environmental noises (snow, Shadow, fog, humility, rain, dew) exist, but in this research out attention is only focused on light intensity because of its common occurrence on a daily basis. Below is the illustration of a light intensity filtering algorithm

Light intensity algorithm has the capability to minimize the effect of light intensity affecting the image [9]. The technique is first carried out on modelling the object reflected by camera based on dichromatic reflection represented as I(x).

Thus, the light intensity detection method is based on the use of dark channel shown in equation (1) and automatic thresholding illustrated in equation (2) is used to label high light reflection in image. This two technique is used for identification for high light area because the intensity value of such are in the dark channel will be having high intensity value while non-area affected with light will be having low intensity value.

$\boldsymbol{I}^{dark}$ of **I** can defined as follows:

$$I^{dark}(X) = \min_{y \in \Omega(x)} \left(\min_{c \in \{r,g,b\}} (I^{c}(y)) \right) \quad (1)$$

$\boldsymbol{\Omega(X)}$ is a local patch cantered at x,x = {x,y} is the image coordinates, $\boldsymbol{I}^{c}$ is a color channel of $\boldsymbol{I}$

$$t^{*} = ArgMax \ \{(1 - p_t)(\omega_1(t)\mu_1^{\,2}(t) + \omega_2(t)\mu_2^{\,2}(t))\} \quad (2)$$

Where t is a threshold value, p is the probability of occurrences at threshold value t, the smaller the Pt is, and the larger the weight will be.

In the mark image generated by automatic thresholding of dark channel image, the area labelled as 1 signifies the area affected with light intensity and 0 signifies area that are not affected with light intensity. Illustration for these expression is given in equation (3).

$$masmask(x) = \{_{0}^{1} \ if \ \boldsymbol{I}(x) > t^{*} \quad (3)$$

In the removal process of Light intensity, Specular to diffuse mechanism is proposed in this technique, illustration

ISSN 1737-9334

of the image with minimal light intensity effect I^D is given in equation (4).

$$I^D(\wedge_{\max}) = I - \frac{\max_{u\in(r,g,b)} I_u - \wedge_{\max} \sum_{u\in(r,g,b)} I_u}{1-3\wedge_{\max}} \quad (4)$$

After filtered images with uncorrupted features generated will now be sent to localization and mapping for further processing of the image which will lead to minimum error in the estimation of the environment.

D. Localization and mapping

The third phase of our technique localization and mapping, the whole procedure and motivation behind simultaneous Localization and Mapping (SLAM) is develop a model of its location and estimate the robot position inside location simultaneously [7]. This cannot be accomplished until the point that robot can refresh its odometry which gives guidance to the robot [15]. Odometry it is continually deceptive in begin of the robot navigation, so the robot needs to explore location in order for vision sensor as an exteroceptive sensor to obtain the scene of the environment in order to redress the robot position [15]. The vision sensor gives scene of the location in form of a picture which needs to go pre-processing phase to filter environmental noise acquired on acquisition phase [15]. The robot would extricate the features from the obtained picture of the earth and re-observing the scene the robot returns to the territory [16].

According to [17] the original SLAM posterior is given as follows:

$$p(x_{1:t}, l_{1:m} | z_{1:t}, u_{o:t-1}) \quad (5)$$

where $x_{1:t}$ represents the path of the robot, $l_{1:m}$ represents the landmarks,

$z_{1:t}$ represents the measurements and

$u_{o:t-1}$ represents the commands given to the robot.

As for static landmarks, the landmark will be defined as follows:

$$l_{1:m} = \{l^s{}_{1:m_s}\} \quad (6)$$

where m represents the number of landmarks, static landmarks. Static landmarks are separated from static as follows. Then (5) is divided into two parts as follows:

$$p(x_{1:t}, l_{1:m} | z_{1:t}, u_{o:t-1})$$

$$= p(x_{1:t}, l^s{}_{1:m_s} | z_{1:t}, u_{o:t-1}).$$

$$= p(x_{1:t}, l^s{}_{1:m_s} | z_{1:t}, u_{o:t-1}) \quad (7)$$

Also Static landmarks can be independent from each other as it's shown below:

$$p(x_{1:t}, l^s{}_{1:m_s} | z_{1:t}, u_{o:t-1})$$

$$= p(x_{1:t}, l^s{}_{1:m_s} | z_{1:t}, u_{o:t-1}). \quad (8)$$

E. Navigation Algorithm

Another fundamental phase in our proposed approach is Navigation phase which is responsible for robot movement around the environment. Navigation is a science of getting mobile robots from place to place [18], the ability of a mobile robot to navigate without external aid [19]. The autonomous navigation of robots in uncontrolled environments is a challenge because it requires a set of subsystems to work together. It requires building a map of the environment, localizing the robot in that map, making a motion plan according to the map, executing that plan with a controller, and other tasks; all at the same time [20]. According to [21] a robot application utilizes path planning and local motion controls to navigate the unknown territory. Path planning studies a model or a map of the location to decide on the regular path points for an autonomous robot to trace from a start site to its destination. Local motion utilizes sensory data to decide a movement that will evade crash with unknown objects or objects whose station in the location had changed [21]. After path planning and local motion, the navigation algorithm will send a command to actuator to instruct the movement of the robot within the unknown location. There are many navigation algorithms which are available from literature each with its strengths and drawbacks however in our method, A* Algorithm is adopted this study to plan a local optimal collision-free path from the current location of the robot to because of its ability to reduces the number of node explorations with respect to Breadth- and Depth-First [20].

F. Dataset

The dataset used was developed from the data used for a PhD Thesis by [23] as cited by [22] on efficient SLAM. "It was documented at the DLR Institute of Robotics and Mechatronics building using a mobile robot controlled by hand. The building covers a region of 60m x 45m and the robot path consists of three large loops within the building (plus a small outside path) with a total length of 505 meters. The robot moved around in the building with artificial landmarks (white/black circles) placed on the ground. The image data has been pre-processed and the relative positions of the observed landmarks with respect to the observation point are provided." [22]. Figure 2 depicts the architectural diagram of the DLR Institute of Robotics and Mechatronics building.

ISSN 1737-9334

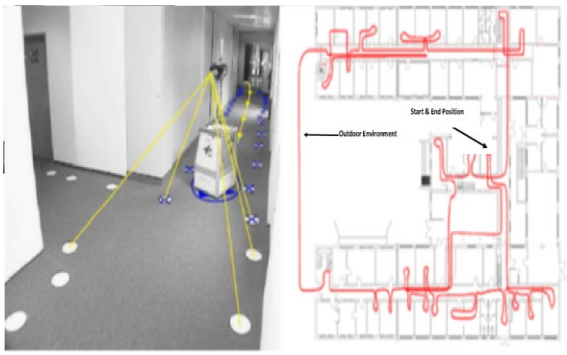


Figure 2: Architectural Diagram of DLR Building [22]

Figure 3 shows a simulated map of the DLR building. The simulated map was also acquired from the study of [22]. The dataset is publicly available, and all the model measurements and set-up of the dataset can be acquired from this "Website, http://www.sfbtr8.spatial-cognition.de/insidedataassociation/, 6 2008" [22].

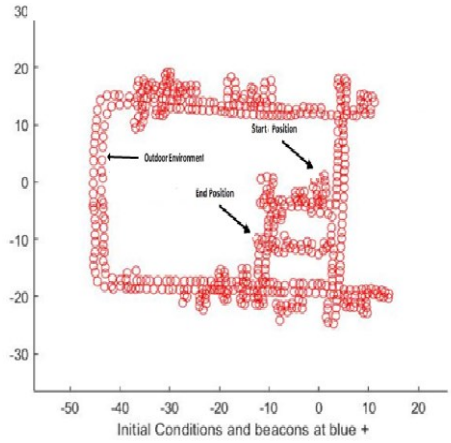


Figure 3: Simulated Map of DLR Building [22]

G. Evalution metrics

The Robot Position of the proposed VSLAM system will be evaluated by employing/calculating the Root mean Square Error (RMSE) and comparing RMSE Robot Position of the proposed system with RMSE of the system from literature. This will be supplemented by calculation of the Robot Position Square Error of the proposed VSLAM system.

The formulation for RMSE is specified in equation (9) as:

$$RMSE = \sqrt{\frac{1}{T}\sum_{K=1}^{T}\left[(X_k - X_k)^2 + (Y_k - Y_k)^2\right]} \quad (9)$$

where $(Y_k - Y_k)$ refers to difference between the actual and $(X_k - X_k)$ refers to estimated distance in Y-direction and refers to difference between the actual and estimated distance in X-direction

The formulation for Square Error at Y-direction is specified in equation (10) as:

$$E^2 = (Y_A - Y_E)^2 \quad (10)$$

where E^2 refers to the calculated square error, Y_A refers to the actual distance in Y-direction, and Y_E refers to estimated distance in Y-direction.

X-direction is specified in equation (11) as:

$$E^2 = (X_A - X_E)^2 \quad (11)$$

where E^2 refers to the calculated square error, X_A refers to the actual distance in X- direction, and X_E refers to estimated distance in X-direction

III. EXPIRIMENTS AND RESULTS

A. Results and Ecaluation Scheme

This chapter defines the experimental methods utilized to evaluate the proposed VSLAM technique that was introduced in the previous chapter. It is imperative to evaluate reliability and viability of the technique and ensure that objectives of the study are accomplished. To validate the proposed VSLAM technique, we compare the proposed VSLAM technique with existing SLAM system result(s) from literature. Utilizing these experimental procedures, the outcomes accomplished are contrasted and examined to perceive how effectively the framework we proposed has improved the VSLAM effectiveness.

Our experimental methods are executed for the determination of error of a proposed VSLAM technique in estimating robot posture in X-direction within the environment and error of a proposed VSLAM technique in estimating the robot posture in Y-direction within the environment.

In uncovering the technique capacity and execution, we evaluated the proposed VSLAM technique using a simulation mathematical software known as Matlab. In comparing a VSLAM technique with existing SLAM systems' results from literature, a study done by [24] on the application of square-root Cubature Kalman filter in SLAM for an underwater robot is utilized to compare for mobile robot position. The next section, Section B, describes the parameters that we are using in our experimental set-up. That is followed by a presentation of simulation results in Section C and then the rest of the chapter presents quantitative results of the simulated VSLAM technique.

B. Expirimental Parametes

In this study, we simulate the proposed VSLAM technique using a vision camera sensor in a pre-defined environment called a DLR building dataset. Table 4-1 presents the parameters employed to simulate the proposed Visual VSLAM technique.

Parameters	**Values**
Robot Speed	0.93 m/s
Maximum Sensor Range	12 m
Camera Field of View	±75°
Image Resolution	320 x 240

Table 1: Parameters

ISSN 1737-9334

C. Simulation

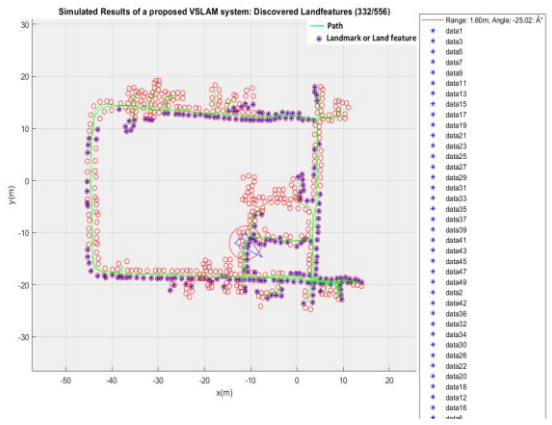


Figure 4-: Simulation Results of a proposed VSLAM system

The simulation map comprises unregularly dispersed landmarks or landfeatures. The mobile robot starts to move at point (0; 0) coordinates and ends at point (-11; -12) coordinates of the simulation map, which is towards middle-left of the simulation map in Figure 4. The mobile robot utilizes the odometry data to predict its movement along the simulation map. In Figure 4, the path of the robot and all the landmarks or landfeatures discovered by the mobile robot are depicted in the map. The light-green line in the simulation map in Figure 4 indicates the path that the robot travels on when navigating the environment. Figure 4 illustrates the linearized least squares result with ground truth of the landmark locations. The Blue stars are the ground truth of landmark or landfeature locations. The ground truth of landmarks and poses of the mobile robot are utilized to update the odometry data as the mobile moves on the simulation environment.

D. Quantitative Analysis of Robot Position

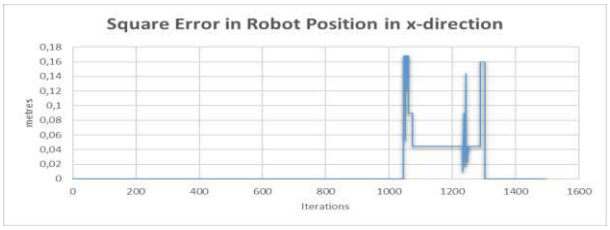


Figure 7: Graphical Results of Error in Robot Position Estimate for X-direction

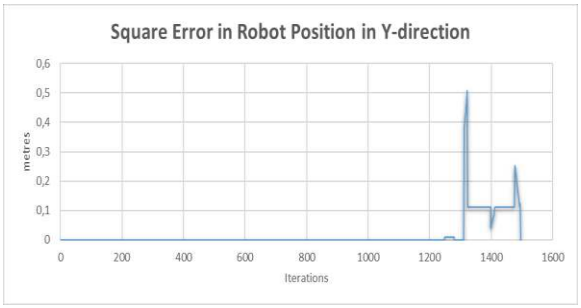


Figure 8: Graphical Results of Error in Robot Position Estimate for Y-direction

Figure 7 and Figure 8 show the amount of square error between the proposed VSLAM technique robot position and the true position within a pre-defined environment in X-direction and Y-direction. Figure 4-3A shows the minimum error of 0 meters and maximum error of 0,168m in X-direction whereas Figure 4-3B shows the minimum error of 0 meters and maximum error of 0,494 meters in Y-direction. These errors are mostly due to large odometry data error and loop closure challenges affected when the robot enters and re-enters the outdoor environment.

RMSE in Robot Position	
SLAM System	**RMSE**
Proposed VSLAM System	0.13 m
Cubature Kalman Filter (CKF) SLAM [24]	0.19 m
Square Root Cubature Kalman Filter (SRCKF) SLAM [24]	0.15 m

Table 2: Comparison of Root Mean Square Error in Robot Position Estimation

Table 2 shows a tabular comparison of a RMSE in robot position of a proposed VSLAM system and SLAM using CKF and SRCKF SLAM algorithm from the study that was done by [24]. Both algorithms, CKF and SRCKF as well as Extended Kalman Filter are extensions of Kalman Filter. Table 4-2 compares RMSE of the two systems. The results on Table 2 show that a proposed VSLAM system RMSE in robot position Estimation is lesser than systems done in the study by [24] simulated using the CKF and Square SRCKF SLAM algorithms. However, both systems utilized different parameters and were tested under different conditions. We can only speculate that our algorithm has a lower RMSE due to the noise filter we used.

IV. CONCLUSION

In this sudy, we have demonstrated the performance of experiments, and the outcomes achieved of the proposed VSLAM technique. After minimising the effects of light intensity by using the Light Filtering Algorithm, the proposed VSLAM system was able to depict a RMSE of 0.13, which is less than the study done by Li et al. (2017). The proposed Visual Simultaneous Localization VSLAM system is able to obtain better results in mapping the location, locating the mobile autonomous robot and navigating the location. Our improved system was able to better classify the position of a robot, discover a satisfactory number of landmarks within a map and determine the true trajectory within the environment.

REFERENCES

[1] S. Khan, D. Wollherr, and M. Buss, “Modeling laser intensities for simultaneous localization and mapping,” IEEE Robotics and Automation Letters, vol. 1, no. 2, pp. 692–699, 2016.

[2] L. Xiang, Z. Ren, M. Ni, and O. C. Jenkins, “Robust graph slam in dynamic environments with moving landmarks,” in Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on. IEEE, 2015, pp. 2543–2549.

[3] J.K. Makhubela, T. Zuva, and O.Y. Agunbiade. "Framework for Visual Simultaneous Localization and Mapping in a Noisy Static Environment." In 2018 International Conference on Intelligent and Innovative Computing Applications (ICONIC).IEEE, 2018 pp. 1-6.

ISSN 1737-9334

[4] C. Park and J.-B. Song, “Illumination change compensation and extraction of corner feature orientation for upward-looking camera-based slam,” in Ubiquitous Robots and Ambient Intelligence (URAI), 2015 12th International Conference on. IEEE, 2015, pp. 224–227.

[5] S. Oh, M. Hahn, and J. Kim, “Simultaneous localization and mapping for mobile robots in dynamic environments,” in Information Science and Applications (ICISA), 2013 International Conference on. IEEE, 2013, pp. 1–4.

[6] B. Clipp, C. Zach, J. Lim, J.-M. Frahm, and M. Pollefeys, “Adaptive, real-time visual simultaneous localization and mapping,” in Proceedings of the 2009 Workshop on Applications in Computer Vision. IEEE, 2009.

[7] X. Gao and T. Zhang, “Robust rgb-d simultaneous localization and map-ping using planar point features,” Robotics and Autonomous Systems, vol. 72, pp. 1–14, 2015.

[8] Davis, J. P., Eisenhardt, K. M., Bingham, C. B. 2007 “Developing theory through simulation methods.” Academy of Management Review, 32:480–499.

[9] Okereke, M., Simeon K.2018 "A Brief Introduction to MATLAB™." *Finite Element Applications*. Springer, Cham, 2018. 27-45.

[10] T. B. Moeslund, Introduction to video and image processing: Building real systems and applications. Springer Science & Business Media, 2012.

[11] N. Shameena and R. Jabbar, “A study of preprocessing and segmentation techniques on cardiac medical images,” 2014.

[12] A. Ballabeni, F. Apollonio, M. Gaiani, and F. Remondino, “Advances in image pre-processing to improve automated 3d reconstruction.” International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 2015.

[13] M. Sonka, V. Hlavac, and R. Boyle, Image processing, analysis, and machine vision. Cengage Learning, 2014.

[14] A. O. Yinka, S. M. Ngwira, Z. Tranos, and P. S. Sengar, “Performance of drivable path detection system of autonomous robots in rain and snow scenario,” in Signal Processing and Integrated Networks (SPIN), 2014 International Conference on. IEEE, 2014, pp. 679–684.

[15] H. Williams, “Human inspired robotic path planning and heterogeneous robotic mapping,” 2016.

[16] F. Heukels, “Simultaneous localization and mapping (slam): towards an autonomous search and rescue aiding drone,” Master’s thesis, University of Twente, 2015.

[17] S. Oh, M. Hahn, and J. Kim, “Simultaneous localization and mapping for mobile robots in dynamic environments,” in Information Science and Applications (ICISA), 2013 International Conference on. IEEE, 2013, pp. 1–4.

[18] U. Frese and G. Hirzinger, “Simultaneous localization and mapping-a discussion,” in Proceedings of the IJCAI Workshop on Reasoning with Uncertainty in Robotics. Seattle, 2001, pp. 17–26.

[19] G. Kosuru, “Design and implementation of an ekf based slam algorithm on a mobile robot,” Ph.D. dissertation, International Institute of Infor-mation Technology Hyderabad-500 032, India, 2011.

[20] E. Z. Gomez,´ “Map-building and planning for autonomous navigation of a mobile robot,” Ph.D. dissertation, Center for Research and Advanced Studies of the National Polytechnic Institute, 2015.

[21] A. Bircher, M. Kamel, K. Alexis, M. Burri, P. Oettershagen, S. Omari, T. Mantel, and R. Siegwart, “Three-dimensional coverage path planning via viewpoint resampling and tour optimization for aerial robots,” Autonomous Robots, vol. 40, no. 6, pp. 1059–1078, 2016.

[22] J. Kurlbaum, U. Frese, “A benchmark data set for data association”. Technical Report, University of Bremen, available online: http://www.sfbtr8.uni-bremen.de/reports.htm Data available on http://radish.sourceforge.net/

[23] U. Frese.2004 “An O(log n) Algorithmm for Simultaneous Localization and Mapping of Mobile Robots in Indoor Environments”. PhD thesis, Technische Fakultat Universit ¨ at Erlangen-N ¨ urnberg,

[24] Li, X., Feng, Y., Huang, R., Zhang, X., Liu, S., & Ai, J. 2017 "The application of square-root cubature Kalman filter in SLAM for underwater robot." *2017 Chinese Automation Congress (CAC)*. IEEE.

ISSN 1737-9334

MODELING OF THE CONCENTRATION EFFECT IN CPV TECHNOLOGIES. CASE OF NOOR OUARZAZATE

Salma El Aimani,

Polydisciplinary of Ouarzazate, Ibn Zohr University, Ouarzazate, Morocco
S.elaimani@uiz.ac.ma

***Abstract*— in this paper, a modelling of a concentrated photovoltaic cell (CPV) is performed. First, we will present generalities and notions on the CPV technology, solar energy and solar radiation. World markets will be cited, but also the Moroccan market with a special mention of Noor Ouarzazate station in southern Morocco. The second part is dedicated to two modelling and simulation of a CPV cell performed under Matlab / Simulink software; these simulations are done under different sunshine conditions, and different concentrations. A comparison between the performances and yields of the two CPV modelling and the PV technology, will be exposed. All for practical validation with real data from the Noor station..**

***Keywords*— PV, CPV, Modeling, Trackers, Irradiation, Matlab**

I. INTRODUCTION

In a difficult energy situation, marked by the foreseeable exhaustion of fossil fuels and their impact on the environment, expectations in terms of renewable energy in general and solar energy in particular, are increasingly important. These energies and in particular solar energy are considered the energy solution of the future. Solar energy is one of the "free" renewable energies, capable of reducing pollution ensuring acceptable performance. The major challenge for researchers and industry in this area is to further improve the efficiency of photovoltaic systems to maximize production facilities and solar power plants. Another major challenge for current research in the field of photovoltaic is lower facilities and power costs through cost reduction systems and photovoltaic devices such as solar cells, modules and photovoltaic panels as well as the mechanical structure to support and position. The cost of photovoltaic cells, which is the most important component of a photovoltaic system as part of energy conversion, remains quite expensive and this is one of the major drawbacks of this energy. Modules or photovoltaic panels and the mechanical structure to support and position is also a big issue in this technology.

The work presented in this paper concerns a modeling of a concentrated photovoltaic cell (CPV). In the first part, we will present generalities on CPV technology, solar energy and solar radiation, and then we describe the main notions of the CPV. World markets will be cited, but also the Moroccan market with a special mention of Noor Ouarzazate station in southern Morocco.

The second part is dedicated to the modelling and simulation of a CPV cell performed under Matlab / Simulink software; these simulations are done under different sunshine conditions, and different concentrations.

In the last part of the paper, a discussion of the simulation results will be presented. All for practical validation with real data from the Noor station.

II. CONCENTRATED PHOTOVOLTAIC (CPV)

A. CPV Technology

Concentrated photovoltaic (CPV) has been widely recognized as the technology that holds the highest hopes for meeting the energy challenges facing the world.

A concentrated photovoltaic system converts light energy into electrical energy in the same way as the conventional photovoltaic technology. The difference with the CPV has the addition of an optical system that concentrates a large area of sunlight onto each cell of the panels, as shown in figure 1. Science is interested in photovoltaic concentrated since the 1970s, but it is only now that the CPV reached commercial viability. This is the latest technology that goes into the solar sector.

CPV systems can be compared with telescopes that follow the sun's position and bring the concentrated light to the cell.

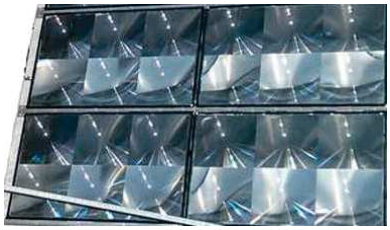


Fig. 1. Concentrating photovoltaic cell

ISSN 1737-9334

B. CPV operating principle

The mirrors focus sunlight onto a small photovoltaic solar cell high efficiency. With this concentration technology, semiconductor materials can be replaced by less expensive optical systems. At equal power, this allows the use of 1000 times less PV material than in the photovoltaic panels to direct sunlight. CPV modules require direct sunlight; it should not be any obstacles like clouds for example, between the sun and modules. This also means that the system must continuously monitor the sun in order to receive direct radiation (or DNI: Direct Normal Irradiation), thanks to a monitoring structure known as "Tracker" because it only works when the sun is visible.

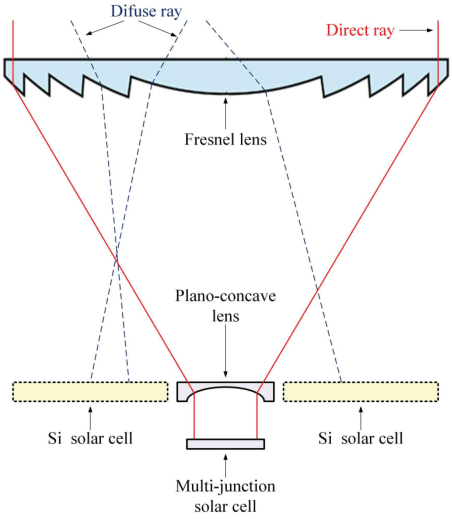


Fig. 2. Fresnel-type CPV System [1]

III. CONCENTRATED PHOTOVOLTAIC TECHNOLOGY (CPV)

C. Parabolic Mirrors

The parabolic mirror reflects the sunlight onto a photovoltaic cell to generate electricity. This technology is also widely used in concentration.

Fig. 3. Parabolic mirrors

The Fresnel mirror is based on the same technology as Fresnel lenses. The mirrors allow flattening a parable in the "cutting" in several sections. The occupied space is small compared to a classic dish and wind resistance is lessened because each mirror movement is controlled by motors. One can thus orient the mirrors to control the concentration (for example to add or remove the focusing of a mirror), or switch to safety position by positioning the flat (in case of strong wind).

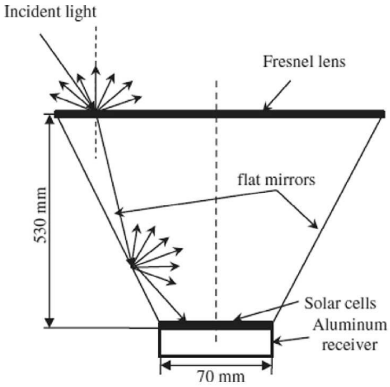


Fig. 4. Fresnel mirror in CPV system [2]

D. Mirrors

Some solar power plants Concentrating (CPV) use for its Fresnel lenses to focus the light rays onto solar cells high efficiency. For the same captured solar flux, energy yields achieved today (30% to 40% for lens and sensor) are double those of the photovoltaic solar panels to direct sunlight (around 10% and up to 20%).

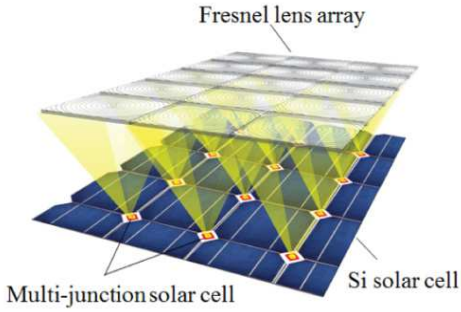


Fig. 5. CPV based on Fresnel lens [3].

IV. E.WOLRD AND MOROCCAN CPV MARKETS

E. World market

CPV technology is subject, like all technologies whose performance affects the economic development, uncertainties and other vicissitudes affecting the world energy market.

In fact, thanks to all these points of strength, CPV technology is booming since 2010 and this has been realized on the market by the appearance of several stakeholders such as module manufacturers and developers of CPV plants tracker. This growth has also led to the installation of large CPV plants of ten megawatt, as shown in figures 6 and 7.

ISSN 1737-9334

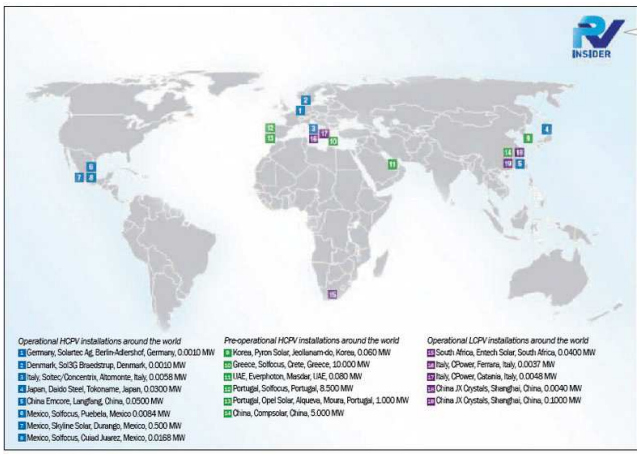


Fig. 6. Largest operating CPV power plants worldwide. [4][16]

Fig 7. Overall forecast of installing the CPV system (megawatts) [5].

F. *In Morocco*

Located in the north-western part of the African Continent, Morocco faces the Sahara Desert in the south and its solar radiation is very high. This means that the country has a great potential for solar energy generation, it stated.

In Morocco, in order to increase its electricity self-sufficiency by making an effective use of its rich solar energy resource, the government intends to introduce 2,000 MW and 4,500 MW solar energy power generation facilities by 2020 and 2030 respectively.

Morocco has a huge solar potential. The country intends to massively exploit this clean and inexhaustible energy in the next decade. The stated objective of the Kingdom is to decrease the external energy dependence from 95% to 85% in 2020. The state has thus set a goal of producing 14% of its electricity needs through solar energy [12].

The year 2018 is a historic year for Noor Ouarzazate. In the first quarter of this year, all plants of this solar complex (Noor Ouarzazate I, II, III and IV) are be in service. And with a capacity of 582 megawatts (MW), these plants will officially Noor Ouarzazate largest multi-technology solar production site in the world. Solar power plants Noor Ouarzazate II, III and IV are in the final stages of construction. They will be ready before the end end of March 2018 [6]. Figure 8 shows the 1MW CPV plant in Ouarzazate at the Noor station.

Fig. 8. CPV plan in Noor IV Ouarzazate

V. MODELLING AND SIMULATION OF CPV CELL

G. *Modelling of CPV cell*

The basic model of a CPV generator is based on the same model of a PV generator except that introduces an optical concentrator (mirror or lens). Figure 9 shows a PV cell.

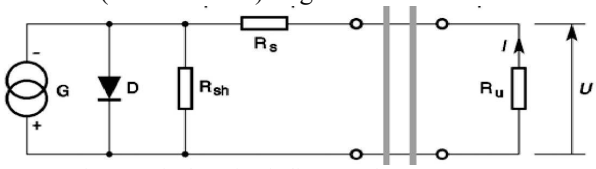


Fig.9. Equivalent circuit diagram of a PV generator [7]

IPH is an ideal current source, D is a diode materializing the fact that the current flows in only one direction. RSH is the shunt resistor that takes into account the inevitable current leaks that occur between the opposite positive and negative terminals of a photocell (micro short circuit in silicon in particular). RS is the series resistance, which is due to the different electrical resistances that the current encounters on its path (intrinsic resistance of the layers, resistance of the contacts). Finally, the load is the impedance of the receiver which imposes the operating point on the photocell as a function of its current-voltage characteristic at the considered irradiation (in the case where the receiver is comparable to a resistor).

n: Ideality factor (practically $1 \leq n \leq 5$). T: Temperature of the junction (in ° K). $I_{PH}=Icc.(E_s/1000)$ Es is the given Irradiation

The CPV cell basic equation is therefore:

$$I = I_{*oc}.\left(\frac{E_s}{1000}\right) - Io.\left(\exp\left(q.\left(\frac{V_{*oc}+R_s.I}{n.K.T}\right)\right) - 1\right) - \left(\frac{V_{*oc}+R_s.I}{R_{sh}}\right) \quad (1)$$

With: $I = I_{ph} - I_d - I_{sh}$

$$I_{ph} = I_{*oc}.\left(\frac{E_s}{1000}\right)$$

$$I_d = Io.\left(\exp\left(q.\left(\frac{V_{*OC}+R_s.I}{n.K.T}\right)\right) - 1\right)$$

$R_s = 0\ \Omega$ (In the ideal case)

$$I_{sh} = \left(\frac{V_{*OC}+R_s.I}{R_{sh}}\right)$$

In the ideal case $R_{sh} = 1M\Omega$

ISSN 1737-9334

$$V *_{oc} = V_{oc} + \frac{K.T}{q.Ln\ C_{opt}}$$

$$I *_{oc} = C_{opt}.I_{sc}$$

With:

I_o = Is: The saturation current of the diode in amperes (A)
I_{Ph}: The photovoltaic current.
q: Charge of an electron (1.6 * 10^{-19} C).
K: Boltzmann Constant (1.38 * 10_{-23} J / K).
n: ideality factor (n=1,5).
T: Temperature of the junction (in k).
V^*_{oc} Open circuit voltage of CPV cell.
V_{oc}: open circuit voltage of PV cell.
I^*_{oc}: The short circuit current of CPV cell.
I_{sc}: The short-circuit current of PV cell ($I_{sc} = 7.34\ A$).
C_{opt}: The optical concentration ratio

H. *Matlab model of CPV cell*

The Matlab model of a CPV cell is shown in figure 10.

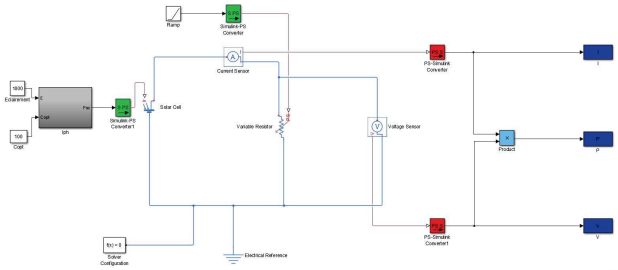


Fig. 10. Model of a CPV cell with PV circuit cell

I. *Simulation results for CPV cell*

Exposed respectively to an irradiation of E=1000W/m² and E=W/m² and for an optical concentration ratio of Copt=100 then Copt=10, the characteristics I =f (V) of the Matlab CPV cell are shown in figure 11.

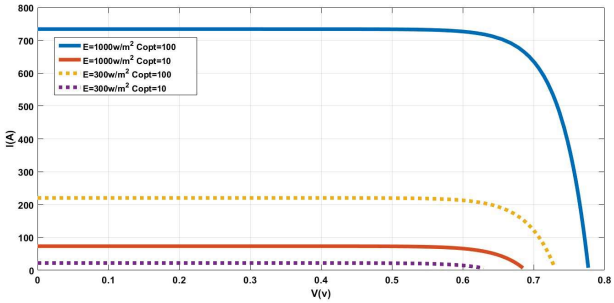


Fig.11. Characteristics I=f(V) of a CPV cell for different irradiations and optical concentration ratios

The characteristics P=f(V) of the Matlab CPV cell under an irradiation of E=1000W/m² and E=300W/m², and an optical concentration ratio of C_{opt}=100 then 10, are shown in figure 12.

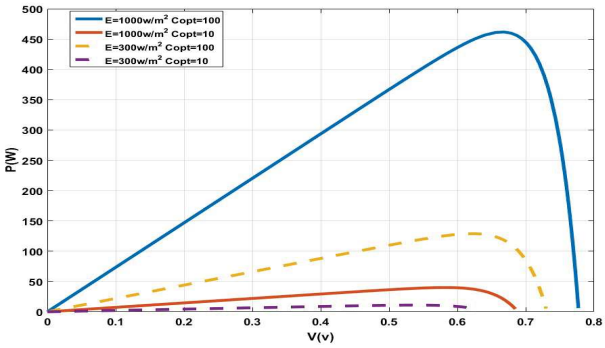


Fig. 12. Characteristics P = f (V) of a CPV cell under different irradiations and optical concentration ratios.

J. *Modelling of CPV cell circuit*

The equation used to simulate the cell CPV circuit is:

$$I - C_{opt}.I_{sc}.\left(\frac{E_S}{1000}\right) - I_O.\left(\exp\left(q.\left(\frac{V_{OC} + R_S.I}{n.K.T}\right)\right) - 1\right) - \left(\frac{V_{OC} + R_S.I}{P_{sh}}\right) \quad (2)$$

The cell simulation diagram CPV circuit is shown in figure 13.

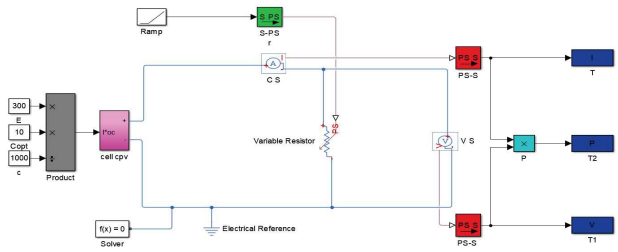


Fig.13. Matlab model of the CPV cell with PV circuit cell.

K. *Simulation results for CPV cell circuit*

Subjected to respectively an irradiation of E=1000W/m², and E=300W/m², and a ratio of optical concentration of Copt =100 then Copt=10, the characteristic I= f (V) of the cell CPV circuit is shown in figure 14.

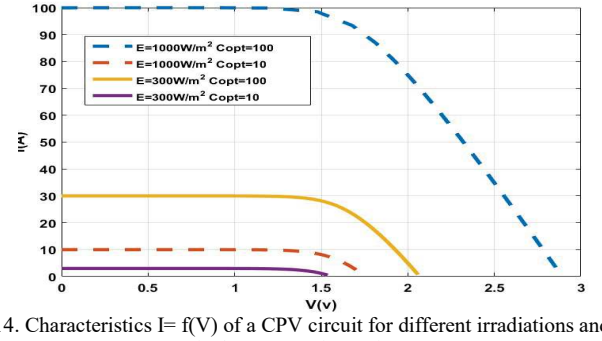


Fig.14. Characteristics I= f(V) of a CPV circuit for different irradiations and optical concentration ratios

The characteristic P = f (V) of the cell CPV circuit under illumination of E=W/m², and E=W/m², and to an optical concentration ratio of Copt= 100 then 10 is shown in figure 15.

ISSN 1737-9334

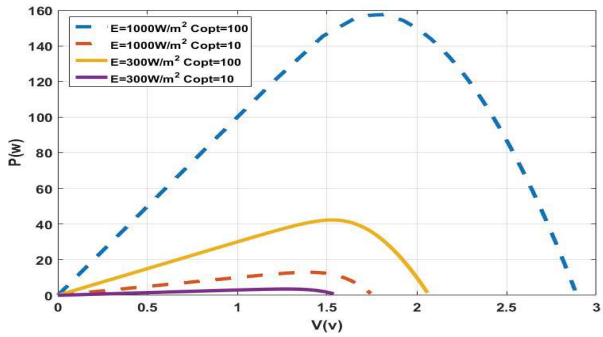


Fig. 15. Characteristics P= f (V) of the cell circuitry CPV circuit under different illumination and optical concentration ratios.

From the simulations above we can conclude that:

- For the voltage: There has been a slight increase in voltage 1.55V between PV [15] and CPV technology.
- For the current: Comparing the simulation results of both PV and CPV cells, there is a current increase of 7.3 A for the PV cell, 90 A as a maximum current value of the CPV cell, with an optical concentration ratio of Copt=10.
- For the electrical power under Copt = 100, there is a large increase in power of 35 W, as the maximum power of the PV cell to the value of 470 W for the CPV cell.

All these comparisons lead us to conclude that the concentration PV cell has a very high energy efficiency compared to a PV cell under the same irradiation. The performance of PV and CPV cell is strongly influenced by climatic conditions, particularly the lighting and the radiation concentration.

VI. CONCLUSIONS

In this paper, we contributed to the comparative study and modelling of CPV technology for power generation,

According to the results, it is good to note that the CPV cells have an important performance versus PV cells, for this reason that we can sum up this performance increase in the concentration effect.

The use of this technology as an emerging renewable energy is highly increasing, thanks to its efficiency, its importance for the preservation of the environment. It is also because it does not emit greenhouse gases, produces no waste and does not involve any significant risk and no significant nuisance.

The energy efficiency is raised when we use the CPV technology compared to the PV one, and under the same environmental conditions. The next step for this work is to implement real data from the Noor CPV station in Ouarzazate, to validate the modelling performed in this study.

REFERENCES

[1] Nguyen Xuan Tien and Seoyong Shin, *A Novel Concentrator Photovoltaic (CPV) System with the Improvement of Irradiance Uniformity and the Capturing of Diffuse Solar Radiation.* Academic Editors: Frede Blaabjerg and Yongheng YangReceived: 16 June 2016; Accepted: 6 September 2016; Published: 8 September 2016.

[2] Noboru Yamada and Kazuya, Okamoto, *Experimental measurements of a prototype high concentration Fresnel lens CPV module for the harvesting of diffuse solar radiation*. Optics Express Vol. 22, Issue S1, pp. A28-A34(2014); •https://doi.org/10.1364/OE.22.000A28.

[3] Xing, Ju Chao, Xu Chao, Xu Zhirong, Liao Zhirong, Liao Yongping ,Yang Yongping, Yang . *A review of concentrated photovoltaic-thermal (CPVT) hybrid solar systems with waste heat recovery* WHR Science Bulletin 62(20). October 2017. DOI: 10.1016/j.scib.2017.10.002.

[4] Aymen Mohamed Sahnoun, "*Contribution to the modelling and high concentration photovoltaic trajectory control Trackers (HCPV)*," Ph.D. HAL Id: tel-01269847https://pastel.archives-ouvertes.fr/tel-01269847. Submitted on 5 Feb 2016.

[5] Letter to Shareholders *Notice of 2014. IHS Annual Report 2013*. Annual Stockholder Meeting. Proxy Statement 2013 Form 10-K Annual Report

[6] Michael Hochberg, *Renewable Energy Growth in Morocco, an Example for the Region*. www.mei.edu. December 2016

[7] Anne Labouret; Pascal Cumunel; John Paul Braun and Benjamin Faraggi, *Solar cells: The basics of photovoltaics*, 5th edition, ETSFs, Paris, (1995).

[8] Christoph Kost, Johannes N. Mayer, Jessica Thomsen,Niklas Hartmann, Charlotte Senkpiel,Simon P. Philipps, Sebastian Nold, Simon Lude, Noha Saad, Jan Schmid, Thomas Schlegl, *Levelized cost of Electricity: PV and PV in comparison to other technologies*. Fraunhofer, ISE (2013).

[9] Yinghao Chu, *Review and Comparison of Different Solar Energy Technologies*, Research Associate, Global Energy Network Institute (GENI), August.

[10] *Photovoltaic guide*. Moroccan Agency for solar Energy: Masen, 2014.

[11] *Current status of concentrator photovoltaic (CPV) Technology.* Fraunhofer Institute for Solar Energy Systems ISE, National Renewable Energy laboratory. CPV Report 1.3, April 2017.

[12] Silvia Martinez Romero, *Concentrating photovoltaic (CPV),* Worldbank, ESMAP, May 6, 2014.

[13] Filipa Reis, *Development of Photovoltaïc systems with concentration*, Phd, Thesis, University of Lisboa. 2013.

[14] Pierre Besson, *Compréhension des comportements électrique et optique des modules photovoltaïques à haute concentration, et développement d'outils de caractérisations adaptés*, January 24, 2017.

[15] Salma El Aimani, *Contribution to a comparative study between PV and CPV technologies. Case of Noor Ouarzazate*, DOI: 10.1145/3234698.3234713, June 2018

[16] Matt Carr *of PV Insider, 3rd Concentrated Photovoltaic Summit* USA, 14-15 November in San Jose, CPV World Map 2011.

ISSN 1737-9334

Energy Performance Analysis of a Solar Chimney Power Plant with and without Thermal Storage System

IKHLEF Khaoula[#1], LARBI Salah[#2]

[#]*École Nationale Polytechnique d'Alger (ENP), Laboratoire de Génie Mécanique et Développement (LGMD)*
10 Avenue Hassen Badi, BP182, El Harrach, Alger, Algérie

[1]khaoula.ikhlef@g.enp.edu.dz

[3]salah.larbi@g.enp.edu.dz

Abstract— **The present work is dedicated to the performance analysis of solar chimney power plants with and without thermal storage system. The performance prediction is carried out according to Schlaich's and Hammadi's mathematical models. The analysis is based on variable solar incident radiation along the day. Obtained results enable us to understand the influence of meteorological conditions, the importance of the thermal storage system and evaluate the effect of geometrical parameters on the power production. Good agreement is observed between the results of this study and those obtained experimentally and theoretically from the literature review.**

Keywords— **Solar Chimney power plant; Energetic performances analysis; Thermal storage; Mathematical model; Sites of Algeria.**

Nomenclature

A_{coll}	Collector area	m^2
Cp_a	Specific heat of air	J/kg K
Cp_s	Specific heat of water-storage	J/kg K
D_{coll}	Solar collector diameter	m
G	Solar intensity	W/m^2
H_{coll}	Height of the collector	m
H_t	Tower height	m
H_s	Water-storage layer thickness	m
h_i	Heat transfer coefficient of inside collector	W/m^2 K
h_∞	Heat transfer coefficient of outside collector	W/m^2 K
k_a	Thermal conductivity of air	W/m^2 K
$\dot{m}_a$	Air mass flow rate	kg/s
t	Time	S
$T_{a,i}$	Air temperature at the collector inlet	K
$T_{a,0}$	Air temperature at the collector outlet	K
T_∞	Ambient temperature	K
T_s	Water-storage temperature	K
$\bar{u}_{coll}$	Average air velocity in the collector	m/s
u_t	Air velocity in the solar tower	m/s
u_{wind}	Wind velocity	m/s
α	Absorptivity of water-storage	/
ρ_a	Air density in the collector	kg/m^3
ρ_s	Water-storage density	kg/m^3
η_{tg}	Turbine-generator efficiency	/
$\Delta\tau$	Day length	hr

I. Introduction

In order to ensure sustainable development and to diversify its energy needs, the world is engaged in an important program of development of renewable energies. To meet its energy needs, it aims to significantly increase the contribution of renewable energies. One of the options that will help meet these demands is the solar chimney power plant (SCPP). The SCPP is a device of renewable energy power plant that transforms solar energy into electricity.

The solar chimney offers a method for large-scale generation of electricity from solar energy. Air is heated near the ground by trapping solar radiation in a flat circular glass-roof greenhouse. The heated air rises in the tower, and the updraft is used to drive a turbine.

The first SCPP prototype was proposed by Schlaich and built in 1982 in Manzanares, Spain [1, 2]. In 1983 Krisst demonstrated a 'back yard type' device with a power output of 10W in West Hartford, Connecticut, USA [3]. In a later study 1984, Haaf reported preliminary test results of the plant built in Spain [2]. In 1985, Kulunk produced a micro scale electric power plant of 0.14 W in Izmit, Turkey [4]. The governing differential equations were developed by Padki and Sherif in 1988 to describe the chimney performance [5]. In 1991 Yan et al. reported on a more comprehensive analytical model in which practical correlations were used to derive equations for the air flow rate, air velocity, power output and thermofluid efficiency [6]. One year later, Padki and Sherif briefly discussed the effects of the geometrical and operating parameters on the chimney performance [7]. In 1997 Kreetz presented a numerical model for the use of water storage in the collector. His calculations showed the possibility of a continuous day and night operation of the solar chimney [8]. In 1999 Bernardes et al. presented a theoretical analysis of a solar chimney operating on natural laminar convection in the steady state [9]. In 2000 Gannon and Backstrom developed an analysis of the solar chimney including chimney friction, exit kinetic losses and a simple model of the solar collector [10, 11]. Kroger and Buys (2001) conducted more thorough analyses of solar chimney power plant performance and Gannon and Von Backstrom (2002)

ISSN 1737-9334

studied the performance of turbines employed in solar chimney power plants [12]. In 2003 Bernardes et al. developed an analytical and numerical model for a solar chimney power plant, comparing simulation predictions to experimental results from the prototype plant at Manzanares [13]. One year later Pastohr et al. conducted a basic CFD analysis on the solar chimney power plant and compared their results to another simple model [14]. In 2005 Schlaich et al. presented the theory, practical experience, and economy of solar chimney power plants to give a guide for the design of 200-MW commercial solar chimney power plant systems [15]. In the same year Bilgen and Rheault developed a mathematical model for evaluating the performance of solar chimney power plants at high latitudes [16]. A refined numerical model for simulating large solar chimney plants was presented by Pretorius and Kroger in 2006 [17]. In 2008 Ming et al. presented a numerical analysis of the flow and heat transfer characteristics in a solar chimney power plant with an energy storage layer [18]. Zhou et al. (2007), Ketlogetswe et al. (2008) and Ferreira et al. (2008) conducted experimental analyses on solar chimney systems [19, 20, 21]. Koonsrisuk and Chitsomboon (2007) and later Zhou et al. (2009) performed numerical simulations of solar chimneys using a commercial CFD software [22, 23]. In 2010 Bernardes et al. evaluated the operational control strategies applicable to solar chimney power plants and Koonsrisuk et al. described the constructal-theory search for the geometry of a solar chimney [24, 25]. So far, some experimental studies have been carried out and several solar chimney pilots in different sizes were constructed Kasaeian et al. (2011), Mehla et al. (2011), Zuo et al. (2012), Kalash et al. (2013), Li and Liu (2014) and Rekaby (2016) [26, 27, 28, 29, 30, 31]. In those researches, the influence of geometrical and climatic parameters on the solar chimney performance was evaluated and temperature distributions in whole system were reported. Siyang and Dennis (2017) to analyse the hydrodynamic features of a series of divergent chimneys in a SCPP [32] established a mathematical model. In the same year, a comprehensive and updated review that includes most of the experimental, analytical and simulation studies, the solar chimney applications, hybrid systems and geographical case studies based on extended references with different focuses in different sections [33]. One year later, the effect of the chimney configuration on the solar chimney power plant performance was investigated by Bouabidi (2018) et al. A series of numerical simulations were conducted to simulate the turbulent flow and an experimental setup was developed in Tunisia to carry out several measurements [34]. Niloufar Fadaeia and Alibakhsh Kasaeian (2018) studied the effect of latent heat storage (LHS) on a solar chimney pilot experimentally [35]. Mathematical models of the solar double-chimney power plant (SDCPP) are established and its performances are analyzed by Fei Cao and his collaborators (2018) [36].

This paper presents a theoretical study of the solar chimney power plant of Manzanares installed in the southern region of Algeria (Adrar). A simplified theoretical model of the solar chimney power plant is described to evaluate the power produced with and without storage system and to estimate the effect of geometrical parameters on the power production of the power plant. The results show that the tower tall, the tower diameter, the collector diameter and the thickness of the water-storage have a significant effect on the power production.

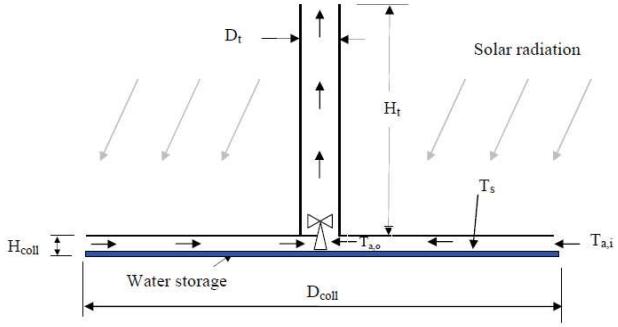


Fig. 1 The solar chimney

II. Mathematical Model

A. Schlaich Model

The schleich model has previously been presented in the article "Technico-economic aspect analysis in the design of solar chimney power plants" [37].

B. Hammadi Model [41]

The simplified heat balance equation of the solar collector shown in Fig. 1 is given as:

$$\alpha\, G\, A_{coll} - h_i\, A_{coll}\,(T_s - T_a) = m_s\, Cp_s \frac{dT_s}{dt} \qquad (1)$$

Where:

$$m_s = \rho_s\, A_{coll}\, H_s \qquad (2)$$

And the energy equation for the air stream through the collector is:

$$h_i\, A_{coll}(T_s - T_a) - h_\infty\, A_{coll}(T_a - T_\infty) = \dot{m}_a Cp_a\left(T_{a,0} - T_{a,i}\right) \qquad (3)$$

Where:

$$T_a = \frac{T_{a,0} + T_{a,i}}{2} \qquad (4)$$

Substitution of equations (2),(3), and (4) into equation(1) gives the following time dependent differential equation:

In replacing $T_{a,0} = 2\ Ta - T_{a,i}$ dans (3), the result obtained with equation (2) is injected into equation (1):

ISSN 1737-9334

$$\frac{dT_s}{dt} = \frac{\alpha G}{\rho_s\, Cp_s\, H_s} - \frac{h_i}{\rho_s\, Cp_s\, H_s}\left[T_s - \frac{\left(\frac{\dot{m}_a Cp_a}{A_{coll}}\right) T_{a,i} + \frac{1}{2}(h_i T_s + h_\infty T_\infty)}{\left(\frac{\dot{m}_a Cp_a}{A_{coll}}\right) + \frac{1}{2}(h_i + h_\infty)}\right] \quad (5)$$

The inside heat transfer coefficient (hi) is taken as [38].

$$h_i = \frac{\left(f/8\right)(Re - 1000)\, Pr}{1 + 12.7\sqrt{f/8}\left(Pr^{2/3} - 1\right)} \frac{k}{D_h} \quad (6)$$

Where:

$$f = [0.79 \ln(Re) - 1.64]^{-2} \quad (7)$$

Where Dh is the hydraulic diameter of the solar collector by considering the flow through the collector as flow between parallel plates of infinite width;

$$D_h = 2H_{coll} \quad (8)$$

And:

$$Re = \frac{\rho_a\, \bar{u}_{coll} D_h}{\mu_a} \quad (9)$$

From the continuity equation:

$$\dot{m}_a = \rho_{a,o} \frac{\pi}{4} D_t^2\, u_t = \rho_a\, \pi\, D_{coll}\, H_{coll}\, u_{coll} \quad (10)$$

Where:

$$\rho_a = \frac{\rho_{a,i} + \rho_{a,o}}{2} \quad (11)$$

The average air velocity through the collector can be expressed as:

$$\bar{u}_{coll} = \frac{\dot{m}_a}{2\,\pi\,\rho_a\,(r_{coll} - r_t)} \int_{r_t}^{r_{coll}} \frac{dr}{r} = \frac{\dot{m}_a}{2\,\pi\,\rho_a\,(r_{coll} - r_t)} \ln\frac{r_{coll}}{r_t} \quad (12)$$

Where r_{coll} and r_t are equal to (D_{coll}/2) and (D_t/2) respectively.

The heat transfer coefficient the collector to the ambient air is given by [38]:

$$h_\infty = 5.7 + 3.8\, u_{vent} \quad (13)$$

- **The Solar Chimney Tower**

The velocity of the hot air at the collector outlet (tower inlet) can be can be estimated using Bernoulli equation as follows:

$$u_t = \sqrt{\frac{2\Delta p}{\rho_{a,o}}} \quad (14)$$

And the pressure difference due to the between the air at the solar tower base and the ambient air is given by:

$$\Delta p = g \int_0^{H_t} \left(\rho_{a,o} - \rho_\infty\right) dH_t = g\left(\rho_{a,o} - \rho_\infty\right) H_t \quad (15)$$

Thus, the equation (14) can be written in term of temperature difference as follows:

$$u_t = \sqrt{\frac{2g\, H_t\left(T_{a,o} - T_\infty\right)}{T_\infty}} \quad (16)$$

The pressure difference is used to accelerate the air and is thus converted to kinetic energy:

$$P_k = \frac{1}{2}\, \dot{m}_a\, u_t^2 \quad (17)$$

The output electrical power of the plant can be found as [39]:

$$P_e = \frac{1}{3} \eta_{tg}\, \rho_{a,o}\, A_t\, u_t^3 \quad (18)$$

The amount of power varies with the variation of incident solar radiation. The equation that describes the amount and variation of solar radiation incident on a clear day is given by the following sinusoidal relation [40]:

$$G = G_g \sin\left(\frac{\pi\, t}{\Delta\tau}\right) \quad (19)$$

Where Gg is the global solar constant which approximately equal to 1000 W/m^2, t=0 for the sunrise and $\Delta\tau$ is the day length which is given by the difference between the sunrise and the sunset.

III. Results and Discussion

The Manzanares solar chimney (Spain, 150km south of Madrid) was used in this study for the simulation with the meteorological data of Adrar. It is a prototype, built between 1982 and 1989 years. The prototype has a tower of 200m high and a collector of 4000m². It reached a production of 44MWh/year, for a peak power of 50kW [41]. Table 1 gives the technical data of Manzanares prototype.

TABLE I
TECHNICAL DATA OF MANZANARES PROTOTYPE [41]

H_T : Tower height [m]	194.6
R_T : Tower radius [m]	5.08
R_C : Mean collector radius [m]	122
H_C : Mean roof height [m]	1.85
U_{vent} : Up wind velocity [m/s]	5
η_e : Turbine efficiency	0.83
α : Friction loss factor	0.9

ISSN 1737-9334

The technical data injected into the Hammadi program, determined according to an average operating temperature of 300K are:

- The water-storage height 10cm;
- The transmittance and absorbance product of the collector 0.65;
- The absorptivity coefficient of storage water 0.8;
- The density of water 1000 kg/m^3;
- The specific heat of water 4178 J/kg K;
- The density of the ground 1900 kg/m^3;
- The specific heat of the ground 840 J/kg K;
- The thermal conductivity of the ground k = 1.26 W/m K;
- The specific heat of the air 1006 J/kg K;

The maximum horizontal solar irradiation and the ambient temperature of the Adrar region (Algeria) are used to analyze the performance of the solar chimney. Meteorological data are taken by METEONORM 7 software with period data (1991-2010). The month of July was chosen for the simulation because it has the highest irradiance, the meteorological data for the Adrar region are illustrated in Fig 2 and Fig 3:

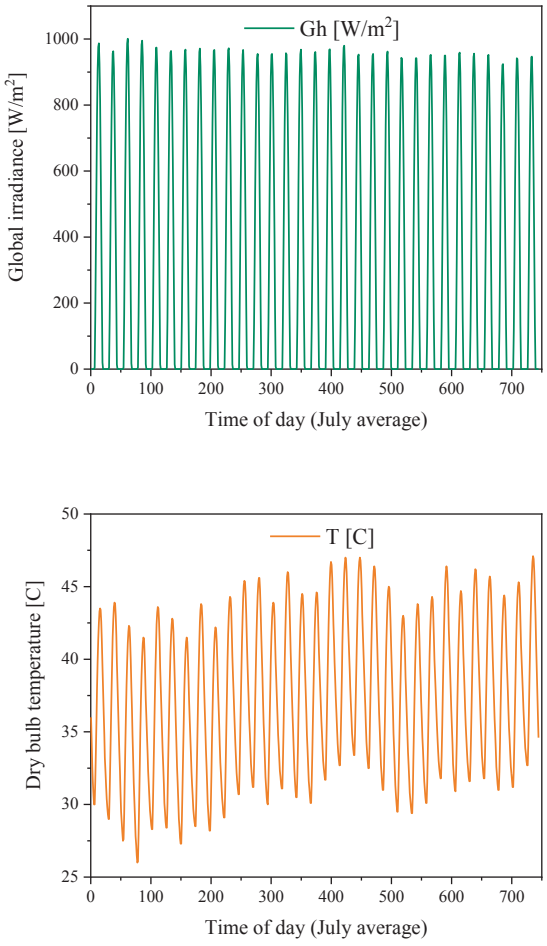


Fig. 2 Hourly average global solar irradiance and temperature for the region of Adrar.

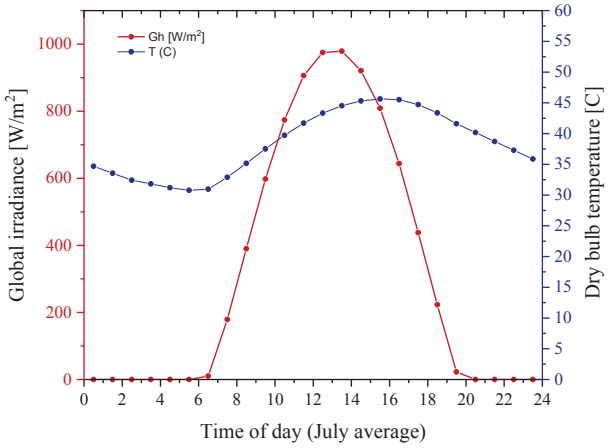


Fig. 3 Daily average global solar irradiance and temperature for the region of Adrar.

In the following analysis, the importance of the power produced by the Manzanares power plant located in the Adrar region with and without a storage system was quantified, and Fig 4 shows the results.

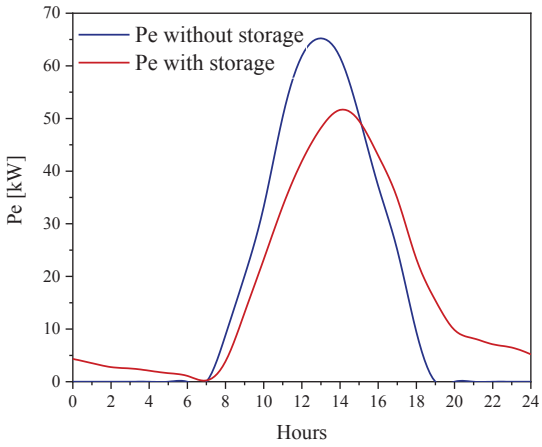


Fig. 4 Power produced with and without storage system (July month)

Figure 4 shows the production of the power plant with and without storage system, the difference can be shown in two ways: a decrease in the peak value and an increase in the min. It can be seen that the power plant with storage system continues to produce energy after the disappearance of the global solar irradiance (null GHI). This generation of energy is provided by the storage system.

In this section, we investigated the effect of the variation of certain technical parameters (geometrical) on the electrical production of the solar chimney power plant equipped with a thermal storage system, by using the meteorological data of the site of Adrar.

ISSN 1737-9334

- Effect of storage height (Hs):

The model used considers the storage medium as a layer located just above the ground and under the collector, with a height Hs and a surface equal to that of the collector. Fig 5 shows the electric power produced for different storage thickness layers such that the storage thicknesses were considered: 5cm, 15cm and 30cm.

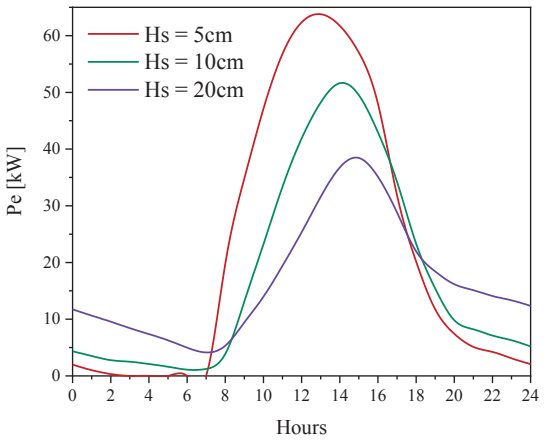


Fig. 5 Effect of storage height on the power production.

Figure 5 shows the effect of water-storage thickness on the output power. This effect can be shown in two ways: a decrease in the peak value and an increase in the min. The more the storage system increases the more the storage effect becomes important at the ends of the curves and therefore the electric power produced during the non-sunny periods is greater.

- Effect of the diameter of the chimney (D_t):

The diameter of the chimney is one of the essential parameters entering into the study of the feasibility in the construction of solar chimney power plant. Fig 6 shows the results of the electric power produced with the variation of the diameter.

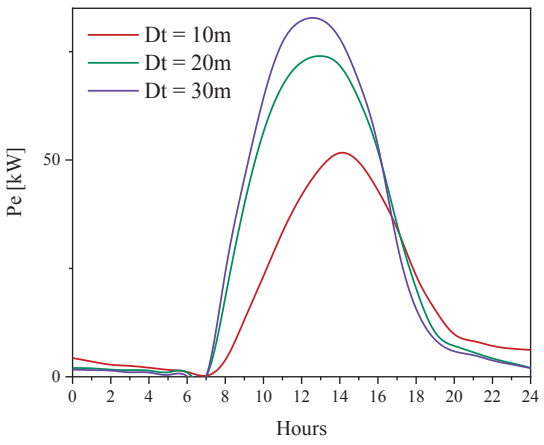


Fig. 6 Effect of the diameter of the chimney on the power production.

Figure 6 shows a significant increase in the power with the tower height due to the increase of the pressure difference between air at the tower base and ambient air as the tower height increases.

- Effect of the height of the chimney (H_t):

The height of the chimney is also considered as one of the parameters entering into the feasibility study of the construction of solar chimney power plant, a very high chimney will require a remarkable investment, an optimization of this parameter is necessary. In this section, we will try to understand the effect of the variation of the height of the chimney on the electricity production of the power plant. Fig 7 shows the results the effect of the variation of the height of the chimney on the power produced.

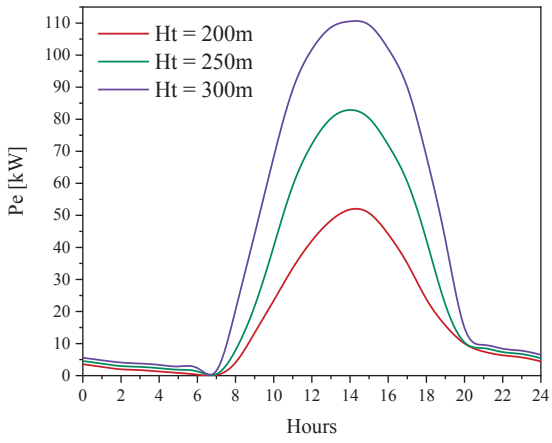


Fig. 7 Effect of the height of the chimney on the power production.

Figure 7 shows that daily electricity production is also directly related to the height of the chimney. It is noted that the higher the height of the chimney, the more the daily electrical power produced increases considerably.

- Effect of collector radius (R_C):

The diameter of the collector represents a fundamental parameter considering the cost necessary for its construction. An optimization of this parameter is necessary to reduce expenses. We will try to understand the effect of the variation of the collector diameter on the electricity production of the solar chimney. Fig 8 shows the effect of collector radius on power produced.

Figure 8 shows that the collector radius has a direct influence on daily electricity production. We notice that the power production of the power plant increases with increasing the collector diameter due to a more solar energy absorbed as the collector area increases.

ISSN 1737-9334

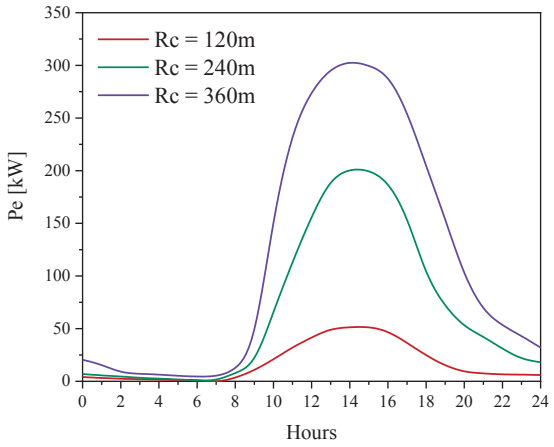


Fig. 8 Effect of collector radius on the power production.

IV. Conclusion

The study presented in this article relates to the analysis of the energy performance of a solar chimney power plant with and without thermal storage support. The site of Adrar was chosen considering the importance of its solar energy potential compared to all regions of Algeria, and the results found shows that:

- The results obtained have shown that thermal storage is the solution to the intermittency of solar radiation such as the storage system increase the power produced during periods of low or no solar irradiance.
- Thermal storage avoids the use of non-renewable energies to guarantee continuous production.
- The optimization study allowed us to understand the variation of the technical parameters on the power produced.

Through the results obtained in this study, it can be noted that a solar chimney plant equipped with a thermal storage system is one of the most interesting alternatives to guarantee a continuous and large-scale energy production. Energy performance can be improved either by improving storage systems by acting on the specific heat of the fluids or solids used in the storage or by acting on the sensible and latent heat storage pairing or the combination of both.

Acknowledgment

I would like to thank my supervisor, Prof. LARBI Salah, for the patient guidance, encouragement and advice he has provided during my thesis in preparing my researches.

References

[1] Haaf W, Fredrich K, Schlaich J. Solar Chimneys, Part I: Principle and Construction of Pilot Plant in Manzanares. J. Solar Energy, Vol.2, 1983, pp.3-20.
[2] Haaf W. Solar Chimneys Part II: Pilot Test Results from the Manzanares Pilot Plant. J. Solar Energy, Vol.2, 1984, pp. 141-161.
[3] Krisst R J K. Energy transfer system. Alternative Source Energy, 1983, 63, 8–11.
[4] Kulunk H. A prototype solar convection chimney operated under Izmit conditions. In: Proceedings of the 7th Miami International Conference on Alternative Energy Sources, Veiroglu, TN, 1985, vol. 162.
[5] Padki M M, Sherif S A. Fluid dynamics of solar chimneys. In: Morrow T B, Marshall L R, Simpson R L. (Eds.). In: Forum on Industrial Applications of Fluid Mechanics, 1988, FED-vol. 70. ASME, New York, pp. 43–46.
[6] Yan M Q, Kridli G T, Sherif S A, Lee S S, Padki M M. Thermo-fluid analysis of solar chimneys. In: Industrial Applications of Fluid Dynamics. ASME, New York, 1991, FED-vol. 132, pp. 125–130.
[7] Padki M M, Sherif S A. A mathematical model for solar chimneys. In: Proceedings of 1992 International Renewable Energy Conference, Amman, Jordan, 1992, vol. 1, pp. 289–294.
[8] Kreetz H. Theoretische Untersuchungen und Auslegung eines tempor. Diplomarbeit, TU Berlin, Berlin, 1997.
[9] Bernardes M A, Dos S, Valle R M, Cortez M F. Numerical analysis of natural laminar convection in a radial solar heater. 1999, International Journal of Thermal Science 38, 42–50.
[10] Gannon A J, Backstrom T W. Compressible flow through solar power plant chimneys. 2000a, Transaction of the ASME Journal of Solar Energy Engineering 122, 138–145.
[11] Gannon A J, Backstrom, T W. Solar chimney cycle analysis with system loss and solar collector performance. 2000b, Transaction of the ASME Journal of Solar Energy Engineering 122, 133–137.
[12] Kroger D G, Buys J D. Performance Evaluation of a Solar Chimney Power Plant. ISES 2001 Solar World Congress. Adelaide, Australia.
[13] Bernardes M A, Dos S, Voss A, Weinrebe G. Thermal and technical analyzes of solar chimneys. 2003, Solar Energy 75, 511–524.
[14] Pastohr H, Kornadt O, Gurlebeck K. Numerical and analytical calculations of the temperature and flow field in the upwind power plant. 2004, International Journal of Energy Research 28, 495–510.
[15] Schlaich J, Bergermann R, Schiel W, Weinrebe G. Design of commercial solar updraft tower systems – utilization of solar induced convective flows for power generation. 2005, Journal of Solar Energy Engineering 127, 117–124.
[16] Bilgen E, Rheault J. Solar chimney power plants for high latitudes. 2005, Solar Energy 79, 449–458.
[17] Pretorius J P, Kroger D G. Solar chimney power plant performance. 2006, Journal of Solar Energy 128 (3), 302–311.
[18] Ming T Z, Liu W, Pan Y, Xu G L. Numerical analysis of flow and heat transfer characteristics in solar chimney power plants with energy storage layer. 2008, Energy Conversion and Management 49, 2872–2879.
[19] Zhou X P, Yang J K, Xiao B, Hou G X. Experimental study of the temperature field in a solar chimney power setup. 2007, Applied Thermal Engineering 27, 2044–2050.
[20] Ketlogetswe C, Fiszdon J K, Seabe O O. Solar chimney power generation project – The case for Botswana. 2008, Renewable and Sustainable Energy Reviews 12, 2005–2012.
[21] Ferreira A G, Maia C B, Cortez M F B, Valle R M. Technical feasibility assessment of a solar chimney for food drying. 2008, Solar Energy 82, 198–205.
[22] Koonsrisuk A, Chitsomboon T. Dynamic similarity in solar chimney modeling. 2007, Solar Energy 81, 1439–1446.
[23] Zhou X P, Yang J K, Wang J B, Xiao B, Hou G X, Wu Y Y. Numerical investigation of a compressible flow through solar chimney. 2009, Heat Transfer Engineering 30, 670–676.
[24] Bernardes M A, Dos S, Backstrom T W. Evaluation of operational control strategies applicable to solar chimney power plants. 2010, Solar Energy 84, 277–288.
[25] Koonsrisuk A, Lorente S, Bejan A. Constructal solar chimney configuration. 2010, International Journal of Heat and Mass Transfer 53, 327–333.

ISSN 1737-9334

[26] Kasaeian AB, Heidari E, Vatan SN. Experimental investigation of climatic effects on the efficiency of a solar chimney pilot power plant. 2011, Renew Sustain Energy Rev 15:5202–6.
[27] Mehla N, Makade R, Thakur NS. Experimental analysis of a velocity field using variable chimney diameter for solar updraft tower. 2011, Int J Eng Sci Technol (IJEST).
[28] Zuo L, Yuan Y, Li Z, Zheng Y. Experimental research on solar chimneys integrated with seawater desalination under practical weather condition. 2012, Desalination; 298:22–33.
[29] Kalash S, Naimeh W, Ajib S. Experimental investigation of the solar collector temperature field of a sloped solar updraft power plant prototype. 2013, Sol Energy; 98:70–7.
[30] Li Y, Liu S. Experimental study on thermal performance of a solar chimney combined with PCM. 2014, Appl Energy; 114:172–8.
[31] Rekaby A. An experimental investigation and CFD analysis of solar chimney power plant. 2016, Ph.D. dissertation, Minia University, Egypt.
[32] Siyang Hu, Dennis Y C. Leung. Mathematical modelling of the performance of a solar chimney power plant with divergent chimneys. 2017, Energy Procedia 110: 440-445.
[33] Kasaeian A B, Molana S H, Rahmani K, Wen D. A review on solar chimney systems. 2017, Renewable and sustainable energy reviews 67: 954-987.
[34] Bouabidi A, Ayadi A, Nasraoui H, Driss Z, Abid M S. Study of solar chimney in Tunisia: Effect of the chimney configurations on the local flow characteristics. 2018, Energy and Buildings 169: 27-38.
[35] Fadaeia N, Kasaeian A. Experimental investigation of solar chimney with phase change material (PCM). 2018, Renewable Energy 123: 26-35.
[36] Cao F, Yang T, Liu Q, Zhu T, Li H, Zhao L. Design and simulation of a solar double-chimney power plant. 2018, Renewable Energy 113: 764-773.
[37] IKHLEF K and LARBI S. Technico-economic aspect analysis in the design of solar chimney power plants. International Conference on Green Energy & Environmental Engineering (GEEE-2018). Proceedings of Engineering and Technology – PET. Vol.37 pp.30-33
[38] Pretorius J, Kroger D G. Critical evaluation of solar chimney power plant performance. J. solar energy, vol.80, 2006, pp. 535-544.
[39] Schlaich J. The solar chimney: electricity from the sun. In: Maurer, editor.
[40] Mawire A, Pherson M. Experimental characterization of a thermal energy storage system using temperature and power controlled charging. J Renewable energy, vpl.33, 2008, pp.682-693.
[41] Hammadi S H. Solar Chimney Power Plant in Basrah. Mechanical Engineering Department- Engineering College University of Basrah.

ISSN 1737-9334

Analysis of Grid Connected Inverter PV-Sytem

Hamed belloumi, Ferid Kourda
Laboratoire Systèmes Electriques (L.S.E.-LR11ES15) Ecole Nationale d'Ingénieurs de Tunis,
Université Tunis El Manar BP.37-1002-Tunis le Belvédère, Tunisie

***Abstract*— This paper focuses on modeling, control, and steady-state of a PV system based on current source inverter. The converter is implemented using a single buck switch, a H-bridge inverter and a CL output filter. The maximum power point is maintained with a Perturb and Observe algorithm. To control the current injected into the grid and the current source inverter output voltage, proportional-resonant controllers are used. Furthermore, to ensure synchronization between the grid current and voltage, a sinusoidal signal generated by a phase-locked-loop is utilized. To evaluate performance of the proposed topology, it has been modeled and simulated in MATLAB/Simulink. The simulation results confirm effectiveness and good performance of proposed converter.**

Keywords—Grid-connected Current Souce Inverter ; Maximum Power Point Tracking (MPPT) ; photovoltaic system;

I. INTRODUCTION

Due to the environmental issues and energy crisis, renewable energy sources have caught the attention of investors and researchers. Among the available renewable sources of energy , the photovoltaic system is considered to be the most promising technology, because of its suitability in distributed generation, satellite systems, and transportation [1]. In distributed generation applications, Photovoltaic systems operate in distributed generation applications in two different modes: grid-connected mode and island mode [2]–[6].

• Standalone: systems where the electricity is supplied to a single charge. This mode is interesting in locations where access to the grid is difficult.
• Connected to the grid: systems where generate electricity is injected into the distribution network.

The majority of new photovoltaic generation systems are connected to the electrical network and enjoys a 70% increase in global capacity in 2008 [2].

The connection between the grid and the PV panels is made through power converters. In the distribution chain, only one or two stages are generally the most used in the photovoltaic converters topologies. In the proposed system a buck and a current source inverter are used.

The conventional voltage source inverter is the most commonly used in grid-connected PV system technology due to its simplicity and availability. The current source converter has not been widely studied for renewable energy systems connected to grid. However, it represents an alternative for the injection of photovoltaic energy into the grid for the following reasons [3]: the input current of the current source converter is smoother than in input current of the conventional voltage source inverter which is preferable for photovoltaic system, and system reliability is increased by replacing the shunt input electrolytic capacitor with a series input inductor.

In the proposed system, P&O algorithm is used to track MPPT. The synchronization of the injected current with the grid is insured by a phase-locked-loop.

The aim of this paper is the analysis of a grid-connected current-source inverter topology for PV applications. To demonstrate the effectiveness and robustness of the proposed system, computer-aided simulation is used to validate the system.

II. SYSTEM DESCRIPTION

The proposed system consists of two stages, as shown in Fig.1. The PV array is connected to a buck converter, in the first stage in order to smoothing the dc link current. We use a sufficiently large input inductor. The second stage is a current source inverter to produce an AC output current to be injected into the grid. Furthermore, a low-pass line L-C filter is used in order to eliminate the switching harmonics.

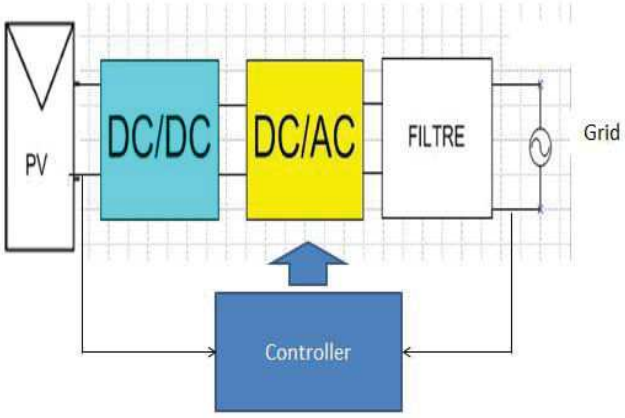


Fig. 1. Proposed current-source inverter topology for the grid-connected PV system.

1. Photovoltaic generator

The photovoltaic panels are composed of panels, wherein the photovoltaic cells are responsible for the conversion of light energy into electricity. The combination of PV modules in series or in parallel can achieve appropriate levels of voltage and current. When combined in series, there will be an increase in the voltage of the PV generator. However, when combined in parallel, there will be an increase in the current generated. Reliable photovoltaic models are essential for simulation of PV power systems to expect energy production from the PV resource under different irradiance and temperature conditions. A solar cell is typically represented by a single diode equivalent circuit. However, PV models can be categorized into main types: a single diode models (SDM) and double diode models (DDM). The DDM; however, its high accuracy it is relatively complex and suffers from low

ISSN 1737-9334

computational speed. The SDM is the most commonly used model in power electronic simulation studies because it offers a tradeoff between simplicity and accuracy [4].
The equivalent circuit of a photovoltaic cell is shown in fig.2. The equivalent current source I_{ph} is the current of the light excitation, while the diode D is the pn junction existing in the photovoltaic cell [5].

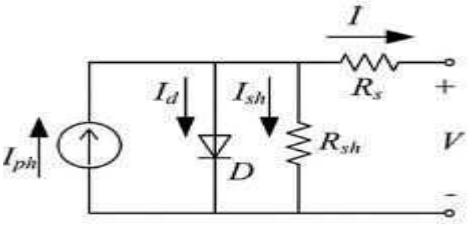


Fig. 2. Equivalent circuits for PV SDM.

Seeking to achieve greater accuracy, the resistance Rsh and Rs are inserted. The shunt resistor R_{sh} represents the difficulty created by the current flow due to crystalline defects (impurities present in the crystalline compound which forms the cell). The series resistor represents the losses associated with pn junction which forms the cell and electrical contact between the terminals and the semiconductor. A detailed study of the effect of this resistance on the performance of PV systems is presented in [6]. The current I_{pv} can be determined by (1). Unfortunately, PV generation systems have two major problems: the conversion efficiency of electric power generation is low (usually less than 17%, especially in low irradiation conditions), and the amount of electric power generated by solar panels changes continuously with the weather [7]. The characteristic of the PV generator is described as follows [8]:

$$I_{pv} = I_{cc} - I_{sat}\left[\exp\left(\frac{(V_{pv} + I_{pv} * R_{ser})}{nV_T} - 1\right) - \left(\frac{V_{pv} + I_{pv} * R_{ser}}{R_{shu}}\right)\right] \quad (1)$$

Where:

$$V_T = \frac{q}{KT} \quad (2)$$

Where
q is the electron charge, K is the Boltzman Constant, T is the temperature, V_{pv} is the output PV voltage, and I_{sat} is the solar cell reverse saturation current.

$$I_{sat} = I_{rr}\left[\frac{T}{T_r}\right]^3 \exp\left[\frac{qE_g}{KA}\left(\frac{1}{T} - \frac{1}{T_r}\right)\right] \quad (3)$$

$$I_{cc} = \left[I_{scr} + K_i(T - T_r)\right]\frac{G}{100} \quad (4)$$

Where K_i is the current temperature coefficient, T_r is the cell reference temperature, G is the solar irradiation, and I_{scr} is the solar cell short-circuit current. Moreover, the cell reverse saturation current is calculated from

$$E_g = E_g(0) - \frac{aT^2}{\mathrm{T}+ß} \quad (5)$$

I_{scr} is the reverse saturation at Tr, E_g is the band gap energy of the semiconductor used in solar cell.
Fig. 3 shows the P-I curve of the PV system.

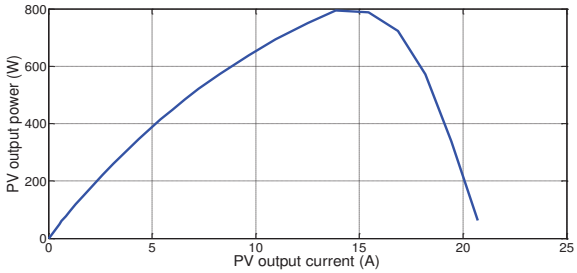


Fig. 3. P-I curve

A. MPPT Algorithm

Due to the non-linearity of the relationship between output parameters of photovoltaic cells - caused by climatic variations (solar radiation and temperature), MPPT techniques are used, which is necessary for maximize the current and voltage generated in a photovoltaic module under any irradiance or temperature condition.
In this method, the instantaneous voltage and current values in the module, V_{PV} (n), I_{PV} (n), are used to calculate the power taken by the photovoltaic module according to this equation [9]:

$$P(n) = V_{PV}(n) * I_{PV}(n). \quad (1)$$

The new power compared with that calculated previously (n-1), where the variation of energy between the two instants of time is extracted.

$$[\Delta P_{PV} = P_{PV}(n) - P_{PV}(n-1)] \quad (2)$$

Positive power variations mean that the disturbance of the voltage of ΔV_{PV} contributes to increase the power, so that the next update ΔV_{PV} will be performed with the same sign.
Once the MPP is found, the next update will send a negative ΔP_{PV}, so the disturbance is the opposite.
However, the P & O technique provides dynamic problems that indicate the size of the time or step used to calculate the duty cycle, which is interpreted in the graphs as oscillations around the operating point. When the disturbance is significant, the system reaches the regime more quickly, but with the highest voltage oscillations all the way to the optimum point.
However, when this step is reduced and the system slows, the module voltage around the MPP varies less [8]. Fig.4 presents this method of Perturb and Observe (P&O):

ISSN 1737-9334

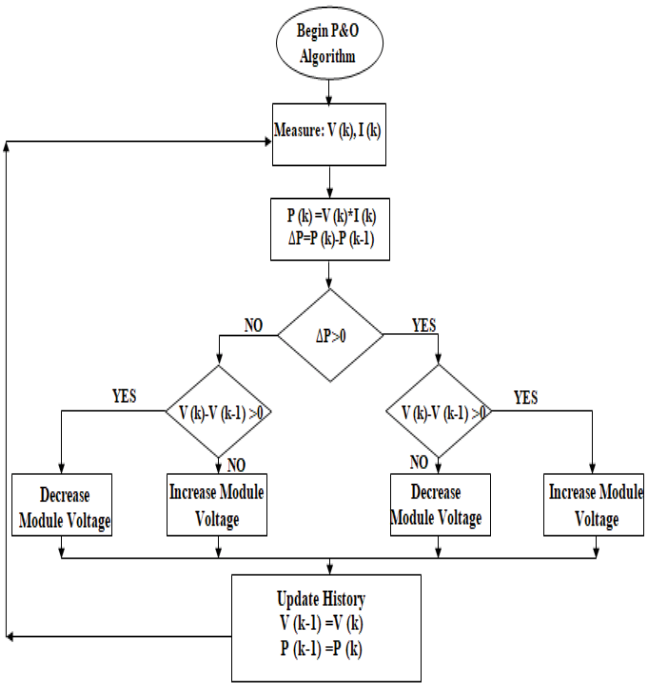


Figure 4: Perturb & Observe algorithm

2. Grid-connected Current Souce Inverter topology

Fig.5 represents a grid-connected PV system using a single-phase current source inverter. The inverter has four insulated-gate bipolar transistors (IGBTs) (K1, K2, K3, K4) and four diodes (D1, D2, D3, D4). Each diode is connected in series with an IGBT switch for reverse blocking capability [2].

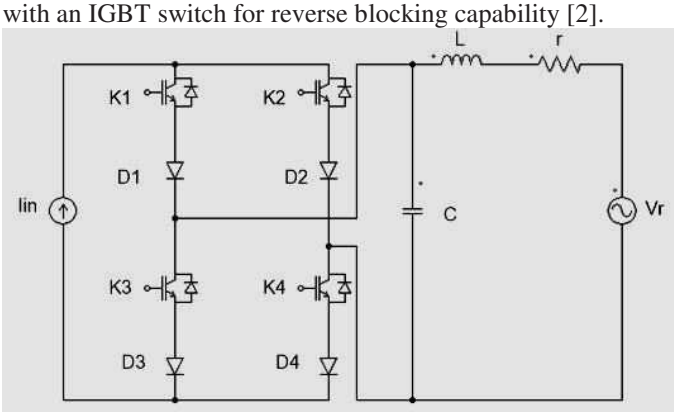


Fig. 5. Single-phase grid connected current source inverter.

Fig. 6. Shows the equivalent circuit of the current source inverter ac side.
The differential equation that describes the ac-side dynamic voltage is:

$$V_c = L\frac{dI_r}{dt} + rI_r + V_r \tag{6}$$

$$C\frac{dV_c}{dt} = d * I_{ond} - I_r \tag{7}$$

From (5) and (6):

$$V_c = \int \frac{1}{C}(d * I_{ond} - I_r) \tag{8}$$

Where d is the duty cycle.
So, the AC output current is defined as following:

$$I_r = \int \frac{1}{L}(V_c - rI_r - V_r) \tag{9}$$

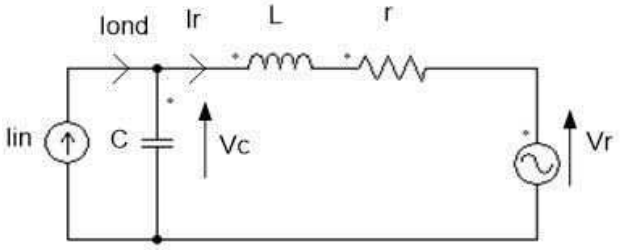


Fig. 6. Equivalent circuit of the current source inverter ac side.

Fig. 7 shows the output current Ir and voltage Vc of the current source inverter.

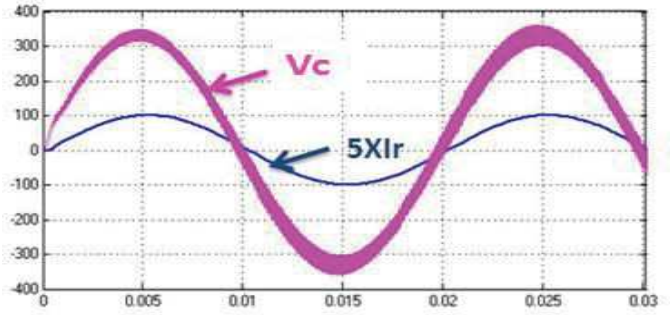


Fig. 7. CSI output current and voltage.

III. COONTROL CURRENT SOURCE INVERTER

To regulate the current Ir and the voltage Vc of the power inverter, the duality between the current source inverter and the voltage source inverter is employed.

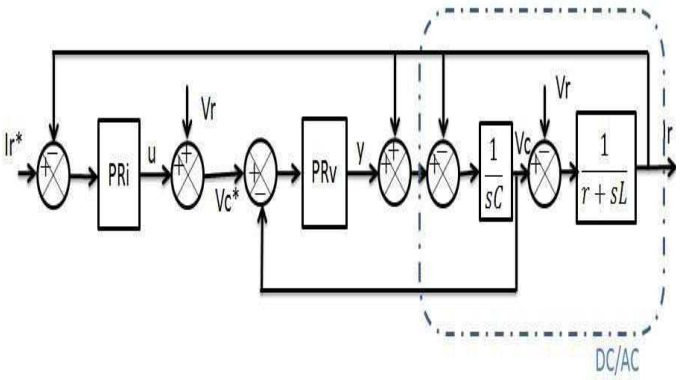


Fig. 8. AC current and voltage loops

ISSN 1737-9334

So, a current control loop (Fig.8) is external and the voltage loop (Fig.9) is internal. The block diagram is shown in Fig.8.
To obtain an accurate control of the current source inverter, proportional resonant (PR) control is used in the voltage and current loop controllers. As mentioned in [13], the PR controllers have better steady state performance in tracking the sinusoidal references and superior transient dynamics than PI controllers.

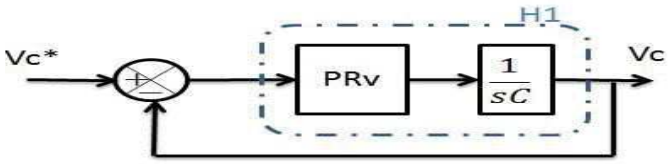


Fig. 9. AC voltage loop.

The PR transfer function of the voltage loop is expressed as:

$$PR_v(s) = \frac{V_c(s)}{V_c^*(s)} = K_{pv} + \frac{K_{iv}s}{s^2+\omega^2} \quad (10)$$

Where K_{pv} is the voltage controller proportional gain and K_{iv} is the voltage controller integral gain.
The transfer function of the voltage loop is defined as following:

$$H_v(s) = \frac{V_c(s)}{V_c^*(s)} = \frac{K_{pv}s^2 + K_{iv}s + K_{pv}\omega^2}{Cs^3 + K_{pv}s^2 + (\omega^2 C + K_{iv})s + K_{pv}\omega^2} \quad (11)$$

As shown in Fig.10, the output voltage Vc is synchronized to the reference Vc*.

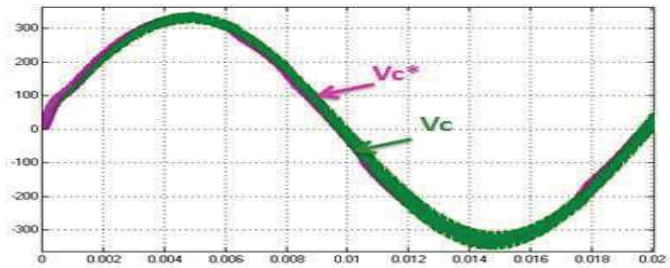


Fig. 10. AC output voltage Vc and reference Vc*.

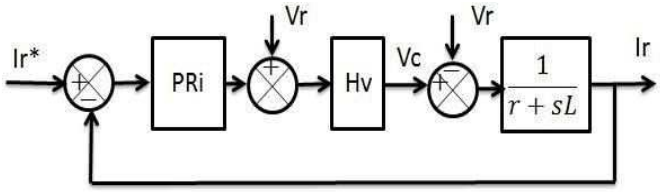


Fig. 11. AC current loop.

The PR transfer function of the current loop is expressed as:

$$PR_i(s) = K_{pi} + \frac{K_{ii}s}{s^2+\omega^2} \quad (12)$$

Where K_{iv} is the voltage controller proportional gain and K_{ii} is the voltage controller integral gain.
The current-loop controller is obtained as:

$$H_i(s) = \frac{I_r(s)}{I_r^*(s)} = \frac{H_v(s)C_{PRi}(s)}{C_{PRi}(s)H_v(s) + Ls + R} \quad (13)$$

As shown in Fig. 12, the output current Ir is synchronized to the reference Ir*.

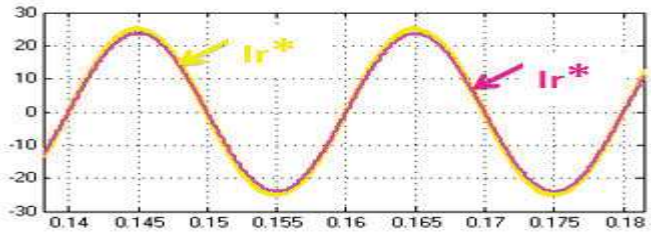


Fig. 12. AC output current Ir and reference Ir*.

a. Phase Locked Loop (PLL)

The quality, reliability and ease of use of the electrical power source are increasingly requested, for this reason the power converters used in grid-connected photovoltaic systems must conform to these requirements.
Since the power factor control is a common objective of these systems requires precise phase information of the mains voltages, named phase tracking system is one of the most important elements.
The phase locked loop (PLL) is the most used technique as a means for synchronizing the phase and frequency of electrical systems. In power electronics, the PLL technique was adopted in the speed control of electric motors.
This is also available to synchronize the voltages and currents applied to connect renewable energy into the network [10].
For these applications, precise monitoring of the voltage vector of the grid is essential to ensure proper operation of the control system. The basic concept of the PLL is that it is a device which causes a signal to follow another. It maintains synchronization of an output signal with an input signal in frequency and phase [11].

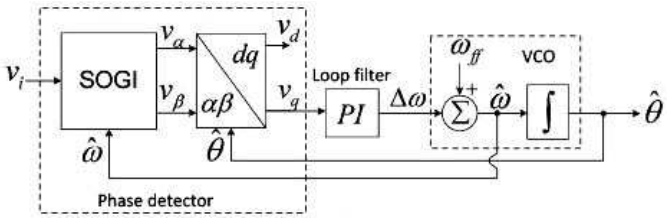


Fig. 13. SOGI-PLL. Basic structure.

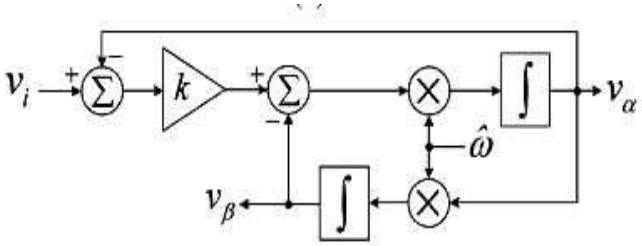


Fig. 14.SOGI block.

Fig.13 shows the structure of the SOGI-PLL, where V_i is the input voltage, $\hat{\omega}$ and $\hat{\theta}$ are the estimated frequency and angle, respectively, and ω_f is the nominal frequency.
The implementation of the SOGI block is shown in Fig. 14 and the Park ($\alpha\beta \rightarrow dq$) transformation is obtained as follows:

$$T=\begin{bmatrix} \cos\hat{\theta} & \sin\hat{\theta} \\ -\sin\hat{\theta} & \cos\hat{\theta} \end{bmatrix} \quad (14)$$

Thus transfer functions of SOGI-PLL can be derived as:

$$G_\alpha(s)=\frac{V_\alpha(s)}{V_i(s)}=\frac{k\hat{\omega}s}{s^2+k\hat{\omega}s+\hat{\omega}^2} \quad (15)$$

$$G_\beta(s)=\frac{V_\beta(s)}{V_i(s)}=\frac{k\hat{\omega}^2}{s^2+k\hat{\omega}s+\hat{\omega}^2} \quad (16)$$

Where k is the damping factor.

IV. Simulation Results

To validate the performance of the proposed system, simulations are performed using MATLAB/Simulink. The design specification and circuit parameters of the proposed system are given in Table I.

TABLE I. CIRCUIT PARAMETERS

Item	Value
DC link inductor (mH)	400
AC line inductor (mH)	2
AC line capacitor (uF)	60
Grid Voltage V rms (V)	220
Grid frequency (Hz)	50
Switching frequency (kHz)	20

To validate the robustness of the algorithm and to search the MPPT, a radiance profile is implemented:
t=[0-0,6s]: is implemented under the test conditions T=25°C and G= 1000*W/m2*
t=[0.6s-0.8s] is implemented under the test conditions T=25°C and G= 800*W/m2*.

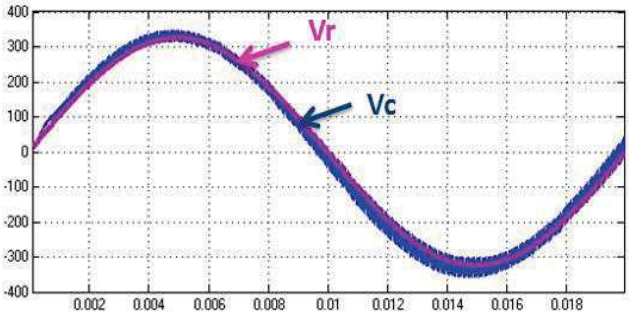


Fig. 14. Synchronization between Vc and Vr.

Fig.14 shows simulated waveforms of the output voltage V_c and the reference voltage V_r, where both are synchronized.
Fig.15 and Fig.16 shows the variation of the output voltage and the output power respectively, corresponding to the irradiance profile, when t=[0-0,6s] the output voltage V_{pv}=57 V and the output power p=800W, which correspond to the coordinates of MPP at T= 25°C and G= 1000 *W/m2*. When t=[6s-8s] the output voltage V_{pv}=50 V and the output power p=600W, which correspond to the coordinates of MPP at T= 25°C and G= 800 *W/m2*.

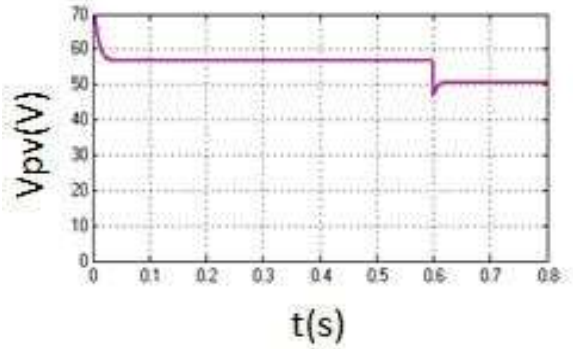


Fig. 16. Variation of the output voltage.

Note that these simulation results show good agreement with the theoretical values.

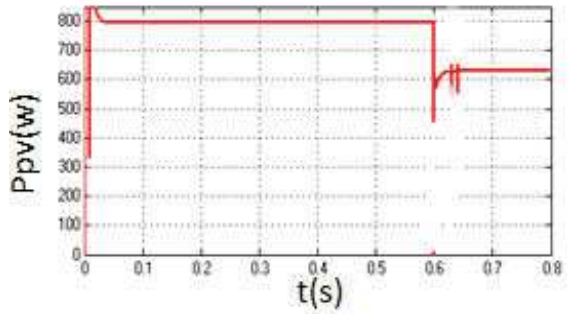


Fig. 16. Variation of the output power.

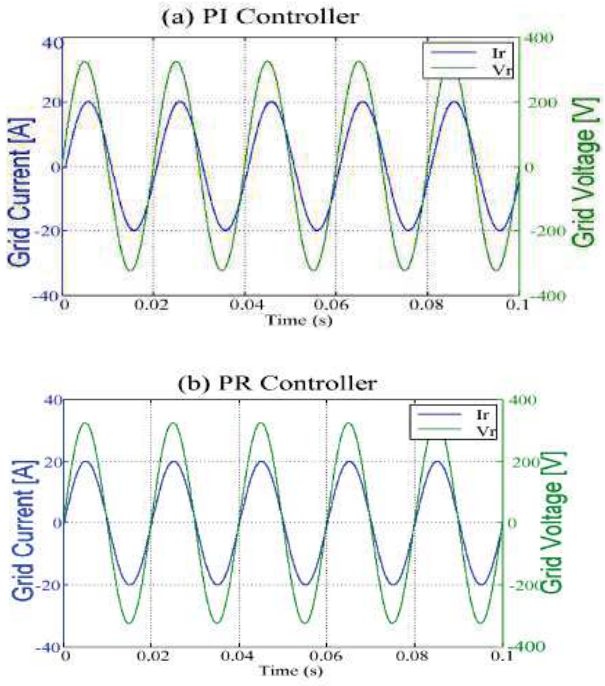
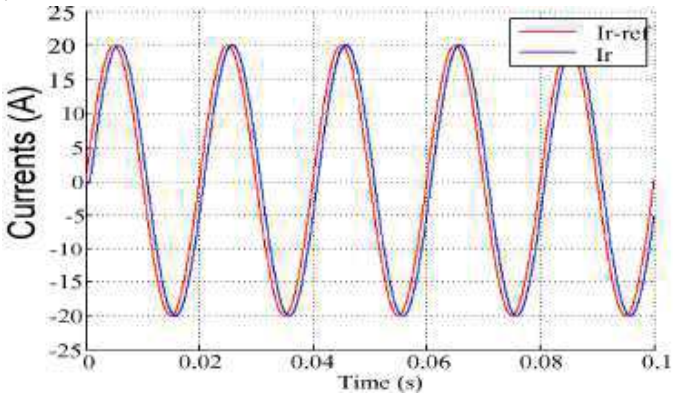


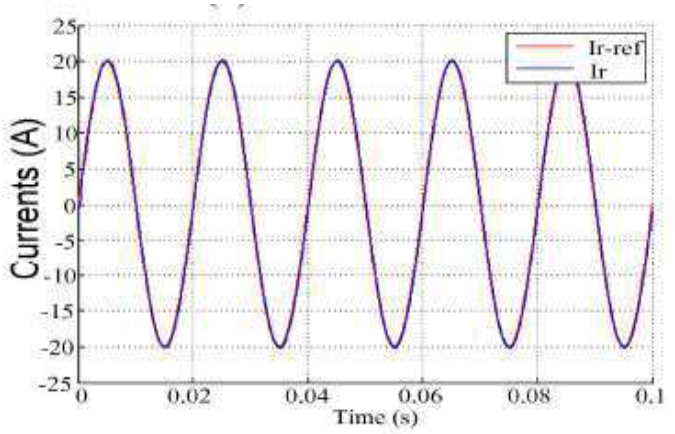
Fig. 17. Simulation results of the proposed grid connected system: (a) PI controller; (b) proposed PR controller

Fig. 17 shows simulation results of the proposed grid connected system with PI controllers and the proposed PR controllers under grid voltage conditions. As shown in Fig. 17(a) the conventional **PI** controllers introduced a phase shift between the grid voltage and the grid current. However, as shown in Fig. 17(b) with the proposed PR controller the current is in phase with the grid voltage.

.



(a) PI controller



(b) proposed PR controller

Fig. 18. Simulation results of the proposed grid connected system: (a) PI controller; (b) proposed PR controller

Fig. 18 shows the simulation responses of the control system using the PR and PI controllers under grid voltage conditions. As shown in Fig. 18(a) the conventional **PI** controllers introduced a phase shift between the current and its reference. However, as shown in Fig. 17(b) with the proposed PR controller the current is perfectly synchronized with its reference and the THD of the grid current is only 0.22%.

V. CONCLUSION

This paper has presented a comparison between standard PI and PR current controllers in A single phase grid connected current source inverter. Simulations using MATLAB/Simulink have been implemented to validate the performance of the proposed system. MPPT, current and voltage loops compound the control structure of the proposed system to insure system performance during normal and varying weather conditions. A SOGI-PLL has been used to guarantee grid requirements. Thus although these results demonstrate the superiority of the PR controller for applications requiring sinusoidal references, additional harmonic compensation is needed in both cases to conform to the standard regulations.

References

[1] S. Dasgupta, S. K. Sahoo, S. K. Panda, and G. A. J. Amaratunga, "Singlephasinverter-control techniques for interfacing renewable energy sources with microgrid—Part II: Series-connected inverter topology to mitigate voltage-related problems along with active power flow control," IEEE Trans. Power Electron., vol. 26, no. 3, pp. 732–746, Mar. 2011.

[2] B. N. AlajmI, Kh.H. Ahmed, G. P. Adam, and B. W. Williams, 'Single-Phase Single-Stage Transformer less Grid-Connected PV System', IEEE *trans. Power electronic.*, VOL. 28, NO. 6, pp 2664-2676, JUNE 2013.

[3] Y Bo, L.Wuhua, Z. Yi, and H. Xiangning, "Design and analysis of a gridconnectedphotovoltaic power system," IEEE Trans. Power Electron.,vol. 25, no. 4, pp. 992–1000, Apr. 2010.

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

[5] Y. Mahmoud and Ehab F. El-Saadany, 'Fast Power-Peaks Estimator for Partially Shaded PV Systems', *IEEE transactions energy conversion* This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[6] E. L. Meyer, E. E. Van Dyk, 'Analysis of the e_ect of parasitic resistances on the performance of photovoltaic modules', Renewable Energy, vol.29 , pp. 333-344, 2004.

[7] V. S. Lacerda, P. G. BARBOSA E Henrique, A. C. BRAGA, _A Single-Phase Single-Stage, High Power Factor Grid-Connected PV System, with Maximum Power Point Tracking', Industrial Technology (ICIT), 2010 IEEE, Pages: 871 _ 877, 2010.

[8] P.P Dash, M. Kazerani, 'Dynamic Modeling and Performance Analysis of a Grid-Connected Current-Source Inverter-Based Photovoltaic System', IEEE *trans. Sust. Energy*, vol. 2, NO. 4, pages 443-450, oct. 2011.

[9] H. Rezk, A.M. Eltamaly "A comprehensive comparison of different MPPT techniques for photovoltaic systems", Solar Energy 112,pp 1–11, 2015.

[10] Se-Kyo Chung, 'A Phase Tracking System for Three Phase Utility Interface Inverters', IEEE *trans. Power Electronic.*, vol. 15, NO. 3, pp 431- 439, MAY 2000.

[1l] S. Jayalath, M. Hanif, "Controller Tuning for a Single Phase Grid Connected Current Source Inverter", ProC. Of fEEE Future E nergy Electronics Conference (lFEEC), pp. 1-6,2015.

[12] C. Hanju, V. Trung-Kien and K. Jae-Eon, "Design and control of PropOltionalResonant controller based Photo voltaic power conditioning system", Proc. OfEnergy Conversion Congress and Exposition 2009. ECCE 2009. IEEE, pp. 2198-2205, 2009.

[13] Guan-Chyun Hsieh and James C. Hung, "Phase-Locked Loop Techniques-A Survey", IEEE trans. indo electronic., vol. 43, no. 6, pp 609-615, Dec. 1996.

[14] S. Golestan, M. Monfared, F.D. Freijedo, and 1. M. guerrero, "Dynamics Assessment of Advanced Single-Phase PLL Structures", fEEE trans. Indo Electronic., vol. 60, no. 6, pp. 2167 - 2177, Jun. 2013.

ISSN 1737-9334

Analysis and Optimization of Planar Transformer for Power Converters Application

Aymen Amouri, Hamed BELLOUMI, Trak BEN SALAH and Férid KOURDA

[#1] *Université de Tunis El Manar, ENIT-L.S.E, BP 37-1002, Tunis le Belvédère, Tunis, Tunisie*
[#2] *Technical and Vocational Training Corporation, College Of Technology in Tabuk, KSA*

***Abstract*— A continuous miniaturization of power converter incorporating planar magnetic components is necessary to higher performance, higher efficiency, higher power and decreasing costs. This paper presents an optimized planar transformer for power converter integration. A comparative study of windings arrangements effects in planar transformer based on Finite Element Method (FEM) for power energy conversion is presented. The most effects for planar transformer design in terms of AC resistance, leakage inductance and parasitic capacitance have been investigated. A planar transformer with three typical winding arrangements structure has been designed and compared to illustrate the effect of the winding configuration. The tradeoffs among these factors have to be analyzed in order to achieve optimal planar transformer design for power converter integration. Simulation results show that full-interleaving winding arrangements can provide good performance on terms of decrease the high frequency copper losses and leakage inductance compared with the other winding arrangements.**

Keywords*— *High frequency planar transformer, leakage inductance, parasitic capacitance, copper loss, magneto motive force (MMF), finite element method (FEM), windings arrangements.

I. INTRODUCTION

In power conversion application, the major trends have been focused on increasing power density, increasing operating frequency, and decreasing profile of power converters [1-5]. Therefore, magnetic components such as inductors and transformers occupied the major space in power systems [5]. The uses of conventional magnetic component structures have many limitations to achieve high performance [6,7]. Therefore, compared with the conventional transformer, the planar transformer has a low profile, higher efficiency, manufacturing repeatability, low electromagnetic interferences and excellent thermal characteristics [3-5]. On the other hand, there are a few aspects which must be improved like decreasing the ohmic loss, core loss, leakage inductance and the parasitic capacitance in planar transformer [2]. Generally, interleaving method between primary and secondary winding of planar transformer is used to decrease leakage inductance and reduce high-frequency ohmic losses [3, 4]. This paper focuses to analysis the parasitic parameters such as AC resistance, leakage inductance and capacitance between winding in planar transformer by analyzing the magneto motive force (MMF) and energy distribution in winding planar transformer. Section II presents the different winding configurations in planar transformer. The magneto motive force (MMF) distribution for each structure is illustrated. The current density distribution at different frequency is presented and analyzed. Based on magneto and electro static simulation, the flux line, magnetic fields and electric fields distribution are analyzed for different windings arrangements. Based on FEM simulation results, a comparative study between winding arrangements in terms of AC resistance, leakage inductance and parasitic capacitance is established and discussed in Section III.

II. PLANAR TRANSFORMER DESIGN DESCRIPTION AND FEM ANALYSIS

Figure 1 presents the PCB planar transformer structure. The construction of such a transformer consists of two main elements: The multi-layer printed on FR4-PCB substrate of primary and secondary windings and planar magnetic core. Each winding in this device is designed by following geometric parameters; the number of layer, the number of turns per layer, the width of the metal line, the spacing between turns, the thickness of metal trace and the thickness of the isolator between layers. A combination of planar EE magnetic core (E43/10/28) is used to design the planar transformer.

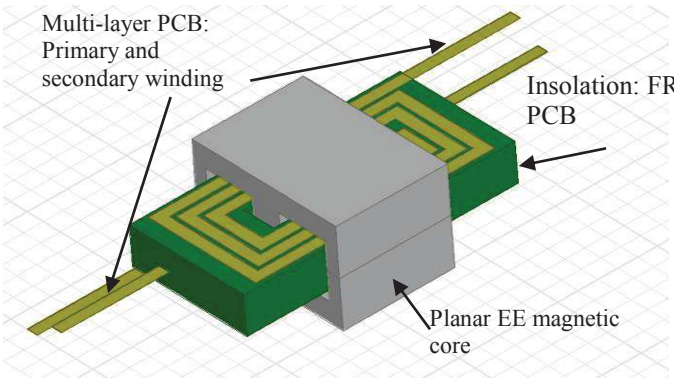


Fig. 1 Prototype of the planar transformer.

For a simple analysis, we choose a planar transformer with two symmetrical windings such way that the

ISSN 1737-9334

number of layers and the number of turns in layers is the same. Primary and secondary winding planar transformer consists of four layers. Each winding layer consists of three turns. For structure 1, the winding layer arrangement is P–P–P–P–S–S–S–S. For structure 2, the winding layer arrangement is P–P–S–S–P–P–S–S. And for structure 3, the winding layer arrangement is P–S–P–S–P–S–P–S. In the three structures, the primary and secondary layers are connected in a series to form a transformer with a 12-to-12 turn ratio. A simplified MMF distributions graph is shown in Figure 2. The arrangement of winding transformer with and without interleaving is the major factor affects the parameters of planar transformer.

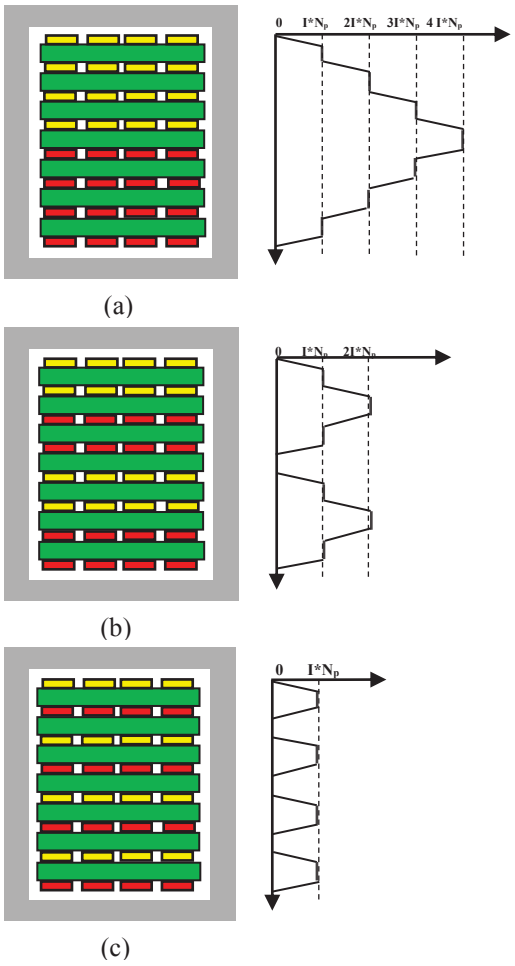


Fig. 2. MMF distributions in winding arrangements :
(a) P-P-P-P-S-S-S-S (b) P-P-S-S-P-P-S-S (c) P-S-P-S-P-S-P-S.

Finite Element Method is used to evaluate the different cases of winding arrangements and to verify the selected design prototype of planar transformer. Figures 3 show the flux lines, the magnetic fields density and electric field distribution for the winding arrangements. Based on eddy current solver, the current density distribution of the different winding arrangements is shown in Figure 4. The analyzed frequency is 50 Hz, 100 kHz and 1 MHz which represents the interesting frequency to illustrate the eddy current effects. At low frequency, the current density distribution is uniform in the cross section of the conductor for different winding arrangements. However, at high frequency, the currents density distribution is concentrated in the surface of conductors. It can be seen that the current distributes more uniformly in the interleaved arrangements than that in non-interleaving winding. This phenomenon is due to the skin and proximity effects in the metal lines [8, 9]. It should be noted that this effect is due to the changes in the magnetic field surrounded by the winding turns [10]. The proximity effect caused by adjacent conductors, effectively increase the skin effect in each individual conductor. Electric field distribution in planar transformer is shown in figure 3. Based on this simulation results, the electrical field is more localized in the FR-4 epoxy insulation between primary and secondary windings. Therefore, a high stray capacitance is obtained when the number of intersections between primary and secondary are increased. Much better results are obtained using the full-interleaving winding design where more uniform magnetic field distributions in the window of planar transformer are acquired.

III. RESULTS DISCUSSION

During the simulation phase, the planar transformer with three different winding arrangements is emulated based on 3D FEM software. Compared with 2D FEM simulation, the 3D FEM takes into account the limited core geometry and its effects on the magnetic structure (e.g. part of the windings no covered by the magnetic core). As an example, three planar transformers designs with different winding arrangements have been designed with EE 43/10/28-3F3 magnetic core materials. The winding arrangements: P-P-P-P-S-S-S-S, P-P-S-S-P-P-S-S and P-S-P-S-P-S-P-S will be analyzed and compared. The letter P and S represents the primary and secondary winding, respectively. These devices are designed using the PCB made on FR-4 material with the dielectric constant, εr, and two PCB substrate thickness 4.4 and 1.0 mm. It is worth mentioning that, the planar transformer parameters are physically-based: the number of turns per layers, the number of layers, the trace width, the thickness of metal trace, the spacing between turns, and the thickness of isolation between layers. Design parameters of the planar transformers are depicted in Table 1. A comparative study between winding configuration on term of the magnetizing inductance, leakage inductance and parasitic capacitance between primary and secondary winding is shown in Table 2.

From figure 4, first, it is noted that for low frequencies the current density is uniform. In fact, resistance is identical to continuous resistance. The resistance factor equal to unity irrespective of the number of layers. In this case the skin depth is much greater than the thickness of the conductors. In second place, when the frequency increases, i.e. the skin depth is comparable to the conductive layer thickness. In fact, the electric current density in the conductor is no longer uniform. Therefore, there was an increase of resistance and especially higher when the number of conductive layers are important.

ISSN 1737-9334

And finally, for higher frequencies, the skin depth becomes significantly smaller than the thickness of a conductive layer. Accordingly, the current density on the device driver, causing a consequent increase in the resistance value. It is important to note that, a demand more effective and commonly is used to interlace the windings, the advantage comes from the fact that halves the maximum value of the field in the winding window. This reduces significantly the leakage inductance and the copper losses of the transformer. However, interleaving is also the disadvantage of increasing the parasitic capacitances.

Parameter symbol	Parameter value	Description
N_p ; N_s	4 ; 4	Number of layers
n_p ; n_s	3 ; 3	Number of turns per layers
w_p (mm)	3	the trace width
S_p (mm)	0.8	spacing between turns
T_p, T_s (mm)	0.1	thickness of metal trace
H_p ,H_s(mm)	1.0	thickness of isolation between layers

nary

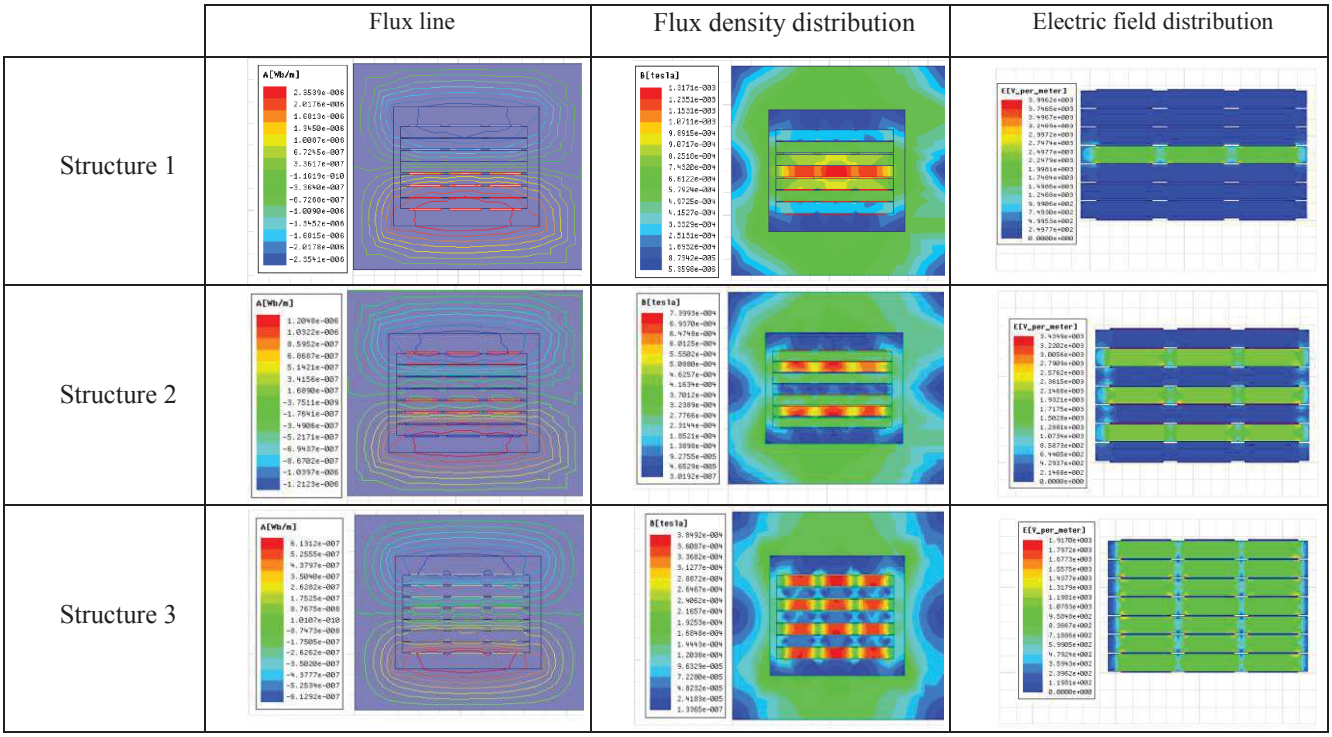


Fig. 3 Comparison of flux line, magnetic field density and electric field distribution of three winding arrangements.

Figure 5 shows a comparative study between different winding arrangements results of the proposed planar transformers. The AC resistance of planar transformers increase with the increased of frequency; this is due to the skin and proximity effects in metal line [4]. On the other hand, the leakage inductance is slightly decreases versus frequency due to the eddy current effects. The skin and proximity effects are responsible for the reduction of leakage inductance [11].

Table 1. Geometrical design planar transformer parameters used during simulation steps.

ISSN 1737-9334

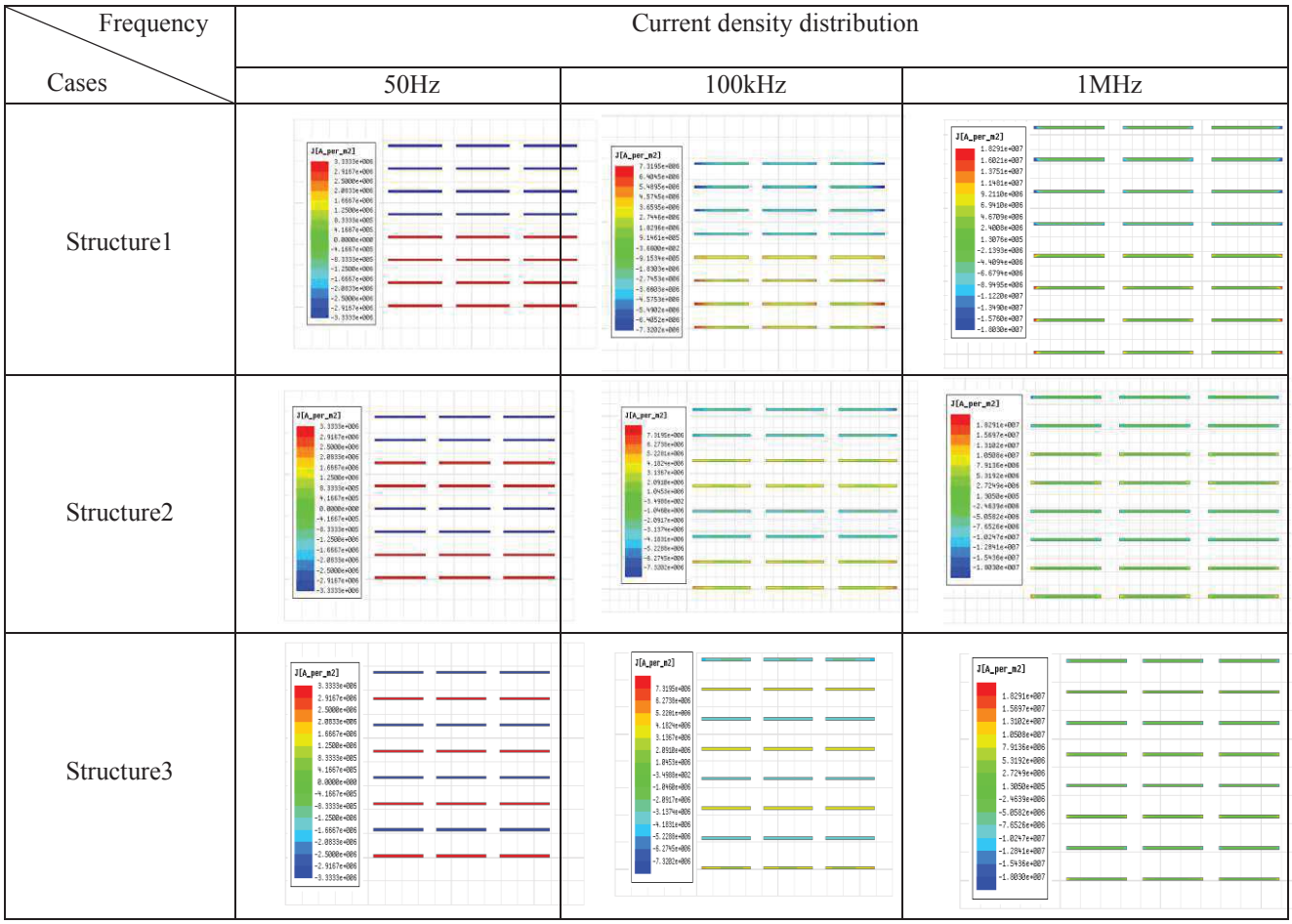


Fig. 4 Current density distribution of three winding arrangements versus frequency.

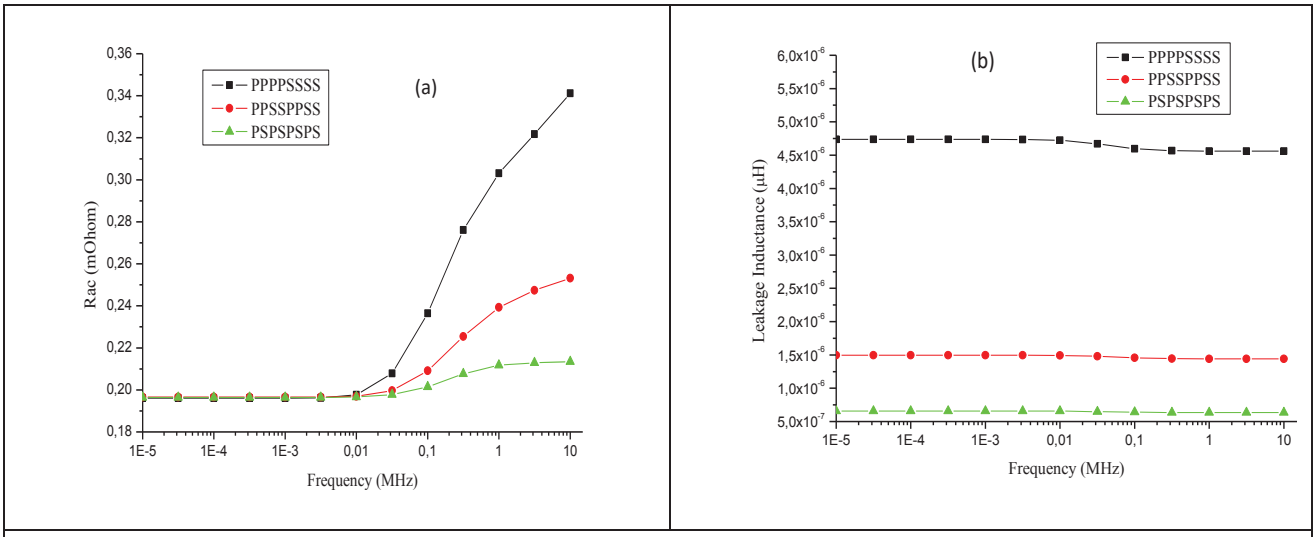


Fig. 5 FEM simulation results for three different winding arrangements of planar transformer: (a) AC resistance (b) Leakage inductance versus frequency.

ISSN 1737-9334

Table 2 illustrates the magnetizing inductance, the simulated leakage inductance and parasitic capacitance between primary and secondary winding for planar transformer based on Magnetostatic and Electrostatic simulation type. It shows that the Noninterleaving has very low parasitic capacitance compared with two other arrangements (Structure 2 and 3). This high capacitance in complete interleaving arrangements for example (Structure 3) is due to the increase intersections between the primary and secondary windings. It is worth mentioning that, the magnetizing inductance in the three arrangements are the same. Comparing the three planar transformer structures, we deduced that the planar transformer with interleaved windings leads to a reduction in copper losses and leakage inductance compared to the transformer, not or partially interleaved. Nevertheless, this technique increases parasitic capacitances between the windings of a planar transformer.

Table 2. Summarized simulation results of different winding configurations of planar transformer: (Magnetostatic and Electrostatic solver type).

	$L_m(\mu H)$	$L_{leak}(\mu H)$	C_{ps} (pF)
PPPP_SSSS	1113.2	4.74	65.2
PP_SS_PP_SS	1111.6	1.5	189.1
P_S_P_S_P_S_P_S	1111.3	0.65	408.8

IV. Conclusion

In this paper a comparative study of windings arrangements effects of HF planar transformer based on Finite Element Method has been presented. A detailed study for designing planar transformer including AC resistance, leakage inductance, and parasitic capacitance between primary and secondary winding has been discussed. Three winding arrangements have been analyzed based FEM simulation and compared to show their advantages and disadvantages in terms of parasitic elements effects. The full-interleaving arrangements structure present that a low winding losses at high frequency, low leakage inductance, and high stray capacitance compared with other winding arrangements. The results of this study can be used to study and design an integrated power converter with high efficiency.

Acknowledgment

This work was supported by the Tunisian Ministry of High Education and Research under Grant LSE-ENIT LR 11ES15.

References

[1] Jianwu Zeng, Wei Qiao, Liyan Qu, and Yanping Jiao,"An Isolated Multiport DC–DC Converter for Simultaneous Power Management of Multiple Different Renewable Energy Sources,"IEEE JOURNAL OF EMERGING AND SELECTED TOPICS IN POWER ELECTRONICS, VOL. 2, NO. 1, MARCH 2014.

[2] Xue Zhang, Wei You, Wei Yao,Shen Chen, Zhengyu Lu, "An Improved Design Method of LLC ResonantConverter," *IEEE 2012*.

[3] Daisuke Tsukiyama, Yasuhiko Fukuda, Shuji Miyake, Saad Mekhilef, Soon-Kurl Kwon, and Mutsuo Nakaoka"A Novel Type High-Efficiency High-Frequency-Linked Full-Bridge DC-DC Converter Operatingunder Secondary-Side Series Resonant Principle for High-Power PV Generation," 2012 International Conference on Renewable Energy Research and Applications (ICRERA), IEEE 2012.

[4] A. K. A. Cho, S. H. Ahn, S. B. Ok, H. J. Ryoo, S. R. Jang, G. H. Rim " Design of LCC Resonant Converter for Renewable *Energy Systems with Wide-Range Input Voltage*," 2012 IEEE 7th International Power Electronics and Motion Control Conference. *IEEE 2012.*

[5] K. Zhang, T. X. Wu, N. Kutkut, J. Shen, D. Woodburn, L. Chow , W. Wu, H. Mustain, and I. Batarseh, "Modeling and Design Optimization of Planar Power Transformer for Aerospace Application," Aerospace & Electronics Conference (NAECON), Proceedings of the IEEE 2009.

[6] Z. Ouyang, O. C. Thomsen and M. A. E. Andersen, "Optimal design and tradeoff analysis of planar transformer in high-power dc–dc converters," IEEE Trans. on Industrial Electronics, vol. 59, no. 7, pp. 2800 - 2810, July 2012.

[7] Jun Zhang, William G. Hurley, Werner H. Wolfle, Maeve C. Duffy,"Optimized Design of LLC Resonant Converters Incorporating Planar Magnetics", IEEE 2013.

[8] Reza Beiranvand, Bizhan Rashidian, Mohammad Reza Zolghadri, and Seyed Mohammad Hossein Alavi,"A Design Procedure for Optimizing the LLC Resonant Converter as a Wide Output Range Voltage Source,"IEEE Trans. Power Electron., vol. 27, no. 8, 2012.

[9] Jun Zhang, Ziwei Ouyang, Maeve C. Duffy, Michael A. E. Andersen, and William Gerard Hurley, "Leakage Inductance Calculation for Planar Transformers With a Magnetic Shunt Transformer," IEEE IEEE Trans. on Industrial Electronics, vol. 50, no. 6, 2014.

[10] Christoph Loef, Rik W. De Doncker, Bernd Ackermann,"On high frequency high voltage generators with planar transformers," Applied Power Electronics Conference and Exposition (APEC), 2014.

[11] Riccardo Pittini, Zhe Zhang, and Michael A.E. Andersen," High current planar transformer for very high efficiency isolated boost dc-dc converters," International Power Electronics Conference (IPEC-Hiroshima 2014 - ECCE-ASIA), 2014.

[12] Wei Chen, Yipeng Yan, Yuequan Hu, and Qing Lu," Model and Design of PCB Parallel Winding for Planar Transformer," IEEE TRANSACTIONS ON MAGNETICS, VOL. 39, NO. 5, SEPTEMBER 2003.

[13] Jun Zhang, William G. Hurley, Werner H. Wolfle, Maeve C. Duffy, "Optimized Design of LLC Resonant Converters Incorporating Planar Magnetics," Applied Power Electronics Conference and Exposition (APEC), 2013 Twenty-Eighth Annual IEEE.

[14] N. Dai, A. W. Lofti, G. Skutt, W. Tabisz and F. C. Lee, "A comparative study of high-frequency, low profile planar transformer technologies,"Applied Power Electronics Conference and Exposition, 1994. APEC '94. Conference Proceedings 1994.

[15] Ziwei Ouyang, Jun Zhang,and William Gerard Hurley, "Calculation of Leakage Inductance for High-Frequency Transformers," IEEE Trans. Power Electron., vol. 30, no. 10, 2015.

ISSN 1737-9334

Comparative study between three active methods of islanding detection with photovoltaic Micro-inverter

Mounir Hlali[1], Imen Bahri [2], Hamed Belloumi[1], and Ferid Kourda[1]
1: Université de Tunis El Manar, ENIT-L.S.E, BP 37-1002, Tunis le Belvédère, Tunis, Tunisie.
2: GeePs | Group of electrical engineering – Paris, UMR CNRS 8507, CentraleSupélec, Univ. Paris-Sud, Univ. Paris Saclay Sorbonne Universités, UPMC Univ. Paris 063 & 11 rue Joliot-Curie, Plateau de Moulon 91192 Gif-sur-Yvette CEDEX, France

***Abstract*-- This article provides an overview of commonly used anti-islanding detection methods, conveniently applied in systems connected to the PV array.**
Anti-islanding methods can generally be classified into three main groups, including passive methods, active methods and basic methods of communication.
Active methods have been the preferred detection technique for years because of the very small non detective zone (NDZ). In this work, three anti-islanding algorithms are evaluated by
Matlab / Simulink® platform with single-stage micro-inverter 300w designed to inject active power into the low-voltage distribution network, using a photovoltaic panel. Normally, SMS method is chosen to make up the anti-islanding detection strategy that goes back to his smallest non detective zone and his simplicity.

Keywords— Grid-connected PV Inverter ;Grid connected mode ; Islanding detection ; island mode ; AFD; SVS; SMS.

I. INTRODUCTION

Renewable energy is defined as the energy that is collected from replenishing resources such as sunlight, wind, ocean waves, geothermal heat etc. Among the renewable energy resources, solar has gained the most attention. Global implementation of photovoltaic has grown exponentially for more than two decades. During this period of time, photovoltaic (PV), also known as solar PV, have evolved from small scale applications to becoming a mainstream electricity source. Based on the International Energy Association [1], solar PV electricity generation will continue to grow rapidly worldwide.
The integration of renewable energy sources and energy storage has raised some concerns for utility power systems. One of the main concerns is islanding. Islanding is a condition where DG system continues to energize the local network, even after getting disconnected from grid [2].
In other words, the system, in addition to supplying electricity to the specified loads, also feeds other loads connected to the network, resulting in operational problems due to the inability of local production, among others even more serious. Although the probability of this effect is extremely low, the standards for the interconnection of photovoltaic systems to the grid require effective methods to detect islanding, such as IEEE 929 2000[3], IEEE 1547[4], and UL1741 [5], due to several resulting problems. Such an islanding can damage the system equipment, affect power system reliability, and endanger the maintenance worker's life.
The importance of islanding detection is highlighted by the fact that many international technical associations such as IEEE or IEC revise and modify the distributed generation (DG) interconnection and islanding codes every two to three years [6]. The power grid can no longer control the voltage and frequency during islanding.
With regards to islanding detection for grid-connected inverters, several methods were studied in literatures [7-8-9]. Normally, islanding detection methods can be classified as communication based methods, passive methods and active methods.
Figure 1 shows a schematic diagram of the Grid connected inverter based DG system. It consists of PV panel, the inverter the local load, the transformer and the utility circuit breaker .

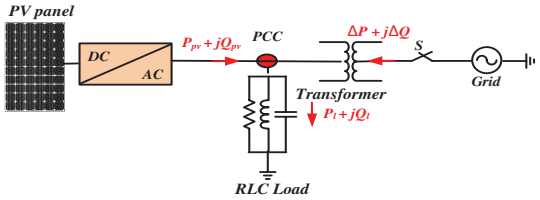


Fig. 1: schematic diagram of the Grid connected inverter based DG system

Passive methods are based on islanding detection through monitoring of parameters such as voltage, current, frequency and/or their characteristics. They interrupt the inverter energy conversion when there are changes beyond some limits previously established, so these methods are not effective in determining the islanding effect correctly, remote detection methods remain economically and technically expensive, and the active methods remain the most effective.In this sense, this article aims to present a study of three active islanding detection algorithms with RLC load, by presenting the control and study strategies.
Several studies were studied the active methods of island detection for connected inverters but they are applied on high power inverters [10], [11],we will present in this paper the application of island detection methods on a micro-inverter with low power 300w.

II. LOAD MODEL

Before defining the different methods of islanding detection, it is important to highlight two key features in order to understand the islanding phenomenon. The first one is associated with the so-called "Non-detection zone" (NDZ), which can be defined as the range (in terms of the difference between the power supplied by the DG inverter and that consumed by the load) in which an islanding detection scheme under test fails to detect this condition. The second one is associated with the nature of loads, which from a conservative point of view can be modeled as a parallel RLC circuit. The reason for using this model is primarily due to the difficulties

ISSN 1737-9334

that some detection techniques have to identify a islanding condition with such loads. Generally, non-linear loads, such loads that produce harmonics or constant power loads do not present difficulties in islanding detection. In particular, RLC loads with high Q factor have problems with island detection. Quality Factor is defined as:

$$Q = R\sqrt{\frac{C}{L}} \quad (1)$$

This parameter describes the relationship between stored and dissipated energy in the RLC circuit. Loads with a high Q have large capacitance and small inductances and/or big parallel resistances.

Before proceeding to the determination of the RLC load parameters, we have chosen to define the usual ZND of a decoupling protection in the plane (ΔQ, ΔP). The photovoltaic inverter system associated with a parallel RLC load shown in Figure1 is quite simple from a theoretical point of view. Indeed, it is possible to adjust the system parameters so that the steady state remains between the voltage and frequency thresholds of the decoupling protection.

P_{PV} and Q_{PV} are the active and reactive powers provided by the photovoltaic inverter. ΔP and ΔQ are the active and reactive powers provided by the grid before opening the switch.

U_{inv} is the voltage in the output of the inverter and the parallel RLC load.

For the static study, it is assumed that the inverter generates constant active and reactive powers. The steady state of such a system is governed by the following equations:

$$U = \sqrt{R\,P_{PV}} \quad (2)$$

$$R\,C\,P_{PV}w^2 + Q_{PV}\,w - \frac{R\,P_{PV}}{L} = 0 \quad (3)$$

Where: ω = 2πf.

The resolution of (3) makes it possible to express the steady-state frequency of the system:

$$f = \frac{-\frac{1}{R}\frac{Q_{PV}}{P_{PV}} + \sqrt{\left(\frac{1}{R}\frac{Q_{PV}}{P_{PV}}\right)^2 + \frac{4C}{L}}}{4\,C\,\pi} \quad (4)$$

If $Q_{PV} = 0$ then the problem is simplified and we notice that the steady-state frequency of the system is exactly the resonance frequency of the parallel RLC load :

$$f = f_{RLC} = \frac{1}{2\,\pi\sqrt{LC}} \quad (5)$$

By fixing P_{PV} and Q_{PV}, The RLC load is considered to consume exactly the P_{PV} and Q_{PV} powers injected by the inverter with nominal voltage and frequency ,the load can be modeled according to [12] - [13]:

$$\begin{cases} U = \sqrt{R\,P_{PV}} \\ Q_f = R\sqrt{\frac{C}{L}} \\ f_{RLC} = \frac{1}{2\,\pi\sqrt{LC}} \end{cases} \longrightarrow \begin{cases} R = \frac{U^2}{P_{PV}} \\ L = \frac{R}{2\,\pi\,f_{RLC}\,Q_f} \\ C = \frac{Q_f}{2\,\pi\,f_{RLC}\,R} \end{cases} \quad (6)$$

In the simulation we fixed the quality factor of the load at $Q_f = 2$ and the load resonance frequency at $f_{RLC} = 50$ Hz.

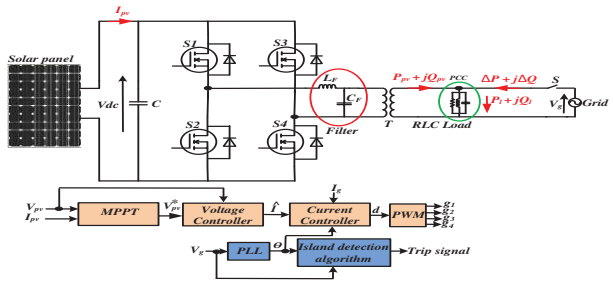


Fig.2: Diagram for testing the islanding detection algorithms

III. ISLANDING DETECTION METHODS

The effects of the disconnection of the network from the operation of the inverter depend on the ratio of the power absorbed by the load and the power available from the PV system. Indeed, the inverter is generally controlled to inject the maximum available energy whatever the load. In case of lack of energy compared to that generated, the network will be responsible for providing the additional need and in case of overshoot, it will absorb. Therefore, if the power generated is greater than the demand, the voltage increases or vice versa.

Similarly, if there is a perfect balance between the load and the generated power, the effect of the network loss becomes invisible for passive islanding methods, as there is no change in the voltage and the output frequency of the inverter.

Figures 3 to 5 illustrate these conditions. It is observed that the worst case for island detection occurs when there is equality between PV production and consumption. In these cases, the Passive methods are not effective; therefore, the active methods are more attractive.

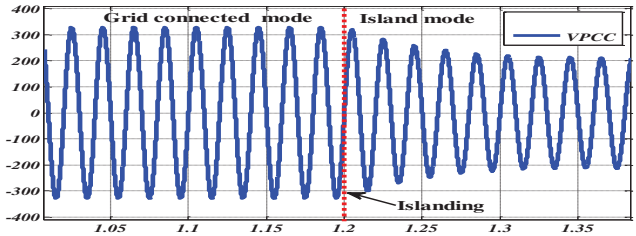


Fig.3: the effect of islanding for a power generated less than the consumption

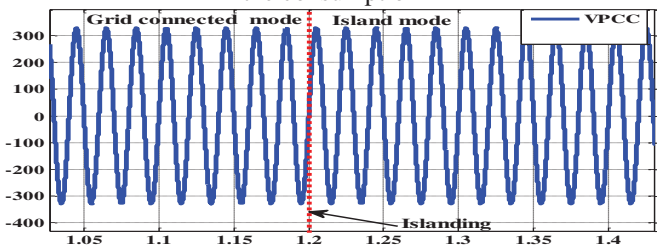


Fig.4 : the effect of islanding for generated power equal to the consumption.

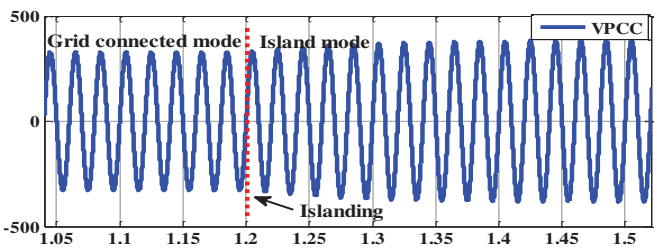


Fig.5: the effect of islanding for generated power greater than to the consumption.

Active islanding detection method is based on the injection of a small disturbance signal to certain parameters at the PCC

ISSN 1737-9334

[13]. The concept of this method is that small disturbance signal will become significant upon entering the islanding mode of operation in order to help the inverter to cease power conversion. Hence, the values of system parameter will be varying during the cessation of power conversion, and by measuring the corresponding system parameters, islanding condition can be detected [10].

Active methods involve feedback control technique that detects changes in the parameters such as frequency or voltage at the PCC. In the case when the PV inverter behaves as a current source. Fig. 6 shows the operation of power flow in active islanding detection. Active methods can effectively reduce, even eliminate, the NDZ and detect islanding accurately. However, this method requires additional control circuits to create adequate disturbances which increase the complexity for implementation. Nevertheless, additional circuits may cause unpredicted effects to the electric power quality, such as the deterioration of the grid voltage quality and system instability [5].

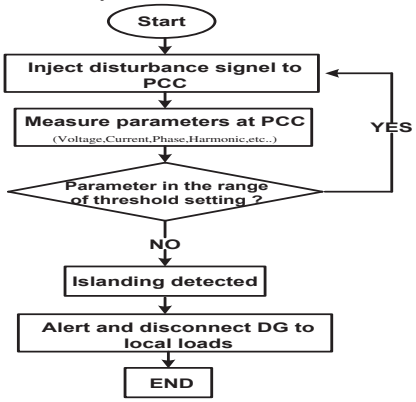


Fig. 6: The flow chart of active islanding detection procedure

a) Active Frequency Drift (AFD)

This method operates with a slightly distorted waveform, often slightly higher than the grid and synchronized with the zero crossing of the same. After crossing the current by zero, it is kept null until it coincides with zero crossing of the mains voltage [13]. In this way, there is a constant tendency of frequency to change under islanding conditions. It is possible to define the chopping factor (C_f) for the AFD method, as a function of the zero current time (Tz) in figure 7 and the period (T) of the grid voltage through (7).

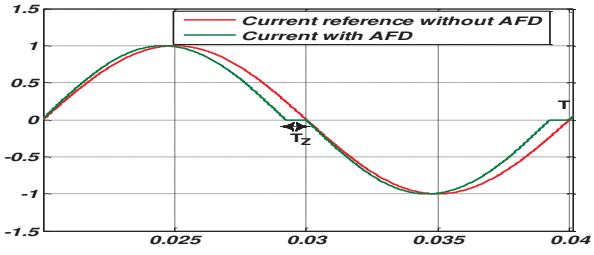


Fig. 7: Inverter current waveform distorted by using AFD

$$C_f = \frac{2\,T_Z}{T} \quad (7)$$

This method can be easily implemented and applied to multiple inverters. However, the AFD method produces a small degradation in the quality of the DG output and the inverter has an NDZ that depends on the value of the chopping factor. The bloc diagram to test this method is represented in figure 8.

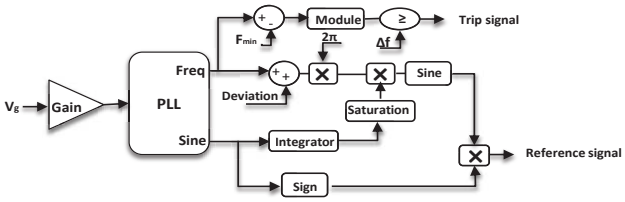


Fig. 8: Strategy for testing the AFD algorithm

b) Sandia Voltage Shift (SVS)

Sandia Voltage Shift (SVS) uses positive feedback technique to prevent islanding based on amplitude of voltage at PCC. When the utility grid is connected, there will be very small or no effect on the power of the system. But once the utility is disconnected, there is reduction in V_{PCC}. According to load impedance's relationship, this reduction will continue and as a result, current and power output reduces. Therefore, this reduction in amplitude of VPCC can be detected by UVP. It is possible either to increase or decrease the power output of the inverter, leading to corresponding OVP/UVP to trip and stops inverter operation [13],[7]. The test algorithm is presented in Figure 9. The algorithm operates in an analogous way for an increase of the effective value of the voltage in the local load.

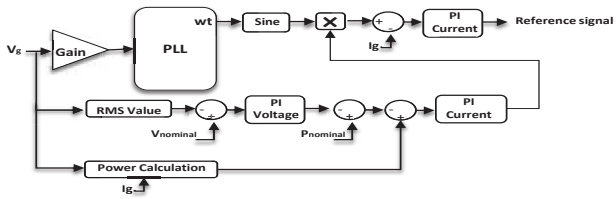


Fig. 9: Strategy for testing the SVS algorithm

c) Slip Mode Frequency Shift (SMS)

The operation principle of the SMS method is based on varying the inverter output frequency by controlling the phase of the inverter current [13],[7]. Usually, DG operates with unity power factor, so in normal operation the inverter output current-voltage phase angle of the inverter, instead to be controlled to be zero, is made to be a function of the frequency of the PCC voltage.

The phase response curve of the inverter is designed such that the phase of the inverter increases faster than the phase of the (RLC) load with a unity-power factor in the region near the grid frequency. This makes the line frequency an unstable operating point for the inverter. While the grid is connected, it stabilizes the operating point at the line frequency by providing a solid phase and frequency reference. However, after the island is formed, the phase-frequency operating point of the load and inverter must be at an intersection of the load line and inverter phase response curve. This method is relatively easy to implement because it is just a slight modification of a component which is already required. Additionally, it has a small NDZ compared with other methods. It also has the advantage of being effective when dealing with multiple inverters and it offers a good compromise between islanding detection, the output power quality and transitory response. However, SMS method requires a decrease in the power quality of the DG inverter.

The bloc diagram to test this method is represented in figure 10.

ISSN 1737-9334

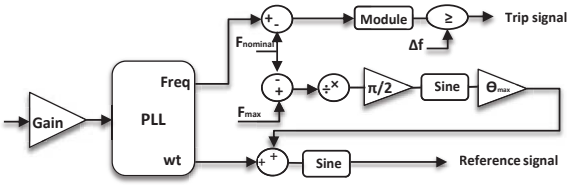


Fig. 10: Strategy for testing the SMS method

IV. SIMULATION AND ANALYSIS OF STUDIED METHODS

Simulation analysis is carried out using MATLAB Simulink. The design specifications and circuit parameters used in simulation are listed in Table I.

Based on the standard IEC 61727 f_{max} = 51Hz, f_{min} = 49Hz, V_{max} = 110% of nominal voltage, V_{min} = 85 % of nominal voltage and Q_f = 2.5 (quality factor) for a fundamental frequency f_0 of 50 Hz and nominal voltage V of 230V

.

TABLE I : PARAMETERS USED IN THE MICRO-INVERTER

Parameter	Value
DC link voltage V_{DC}	35[V]
Power of PV system	300[W]
Quality factor	2.5
Resistance of inverter load	176.33[Ω]
Inductance of inverter load	28.07[mH]
Capacitance of inverter load	361.2[µF]
Resistance of the grid	0.2[Ω]
Inductance of the grid	0.2e-3[H]
RMS voltage	230[V]

a. Active Frequency Drift (AFD)

b.

A variation of 5 Hz was specified for the frequency of the current injected into the grid, corresponding to a fraction of 10 % of the grid period. After disconnecting the network, the frequency of the local load voltage increases, as shown in Figure 11,figure12 presented the voltage V_{PCC} after islanding , from this figure we notice that the voltage decreases also.

Figure 13 shows the total harmonic distortion of current injected by the inverter to the RLC load after the application of the AFD method we note that after the application of this method the THD is of the order of 11.5%.

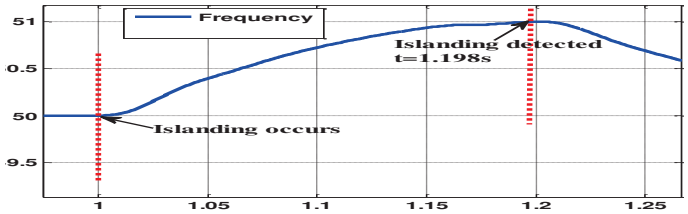


Fig.11: Frequency change after disconnection from the grid – Method: AFD.

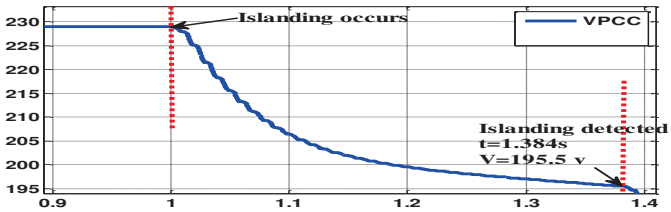


Fig.12: Voltage at PCC change after disconnection from the grid – Method: AFD

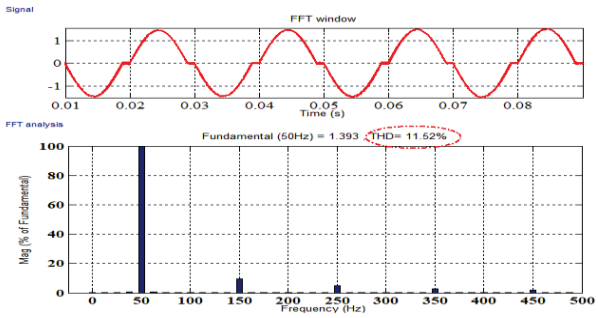


Fig.13: The THD of the output current of the inverter – Method: AFD

We can notice a short time detection, islanding occurs at t=1s and the islanding detected at t=1.198s so the time detection is Δt=198ms , the islanding detected with frequency and not the case for the voltage which is in the non-detective zone at t=1.198s.

c. Sandia Voltage Shift (SVS)

The specification for detection by the SVS was 85% of the nominal voltage and the time for detection can be visualized in Figure 14, where after 1s the network is disconnected and the algorithm starts to reduce the injected power until it reaches the minimum voltage.

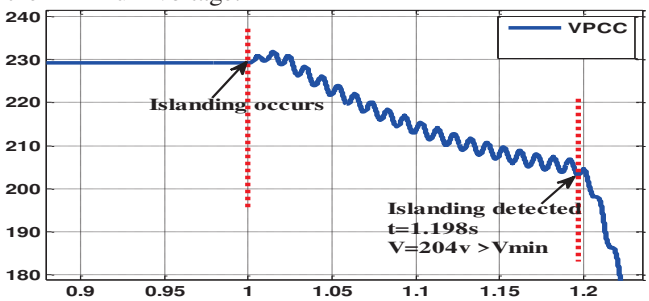


Fig.14: Voltage at PCC change after disconnection from the grid – Method: SVS

Figure 15 shows the total harmonic distortion of current injected by the inverter to the RLC load after the application of the SVS method we note that after the application of this method the THD is of the order of 3.9%.

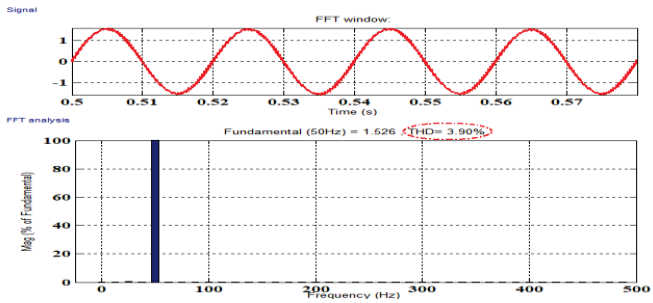


Fig.15: The THD of the output current of the inverter – Method: SVS

We can notice from figure 14 a short time detection, islanding occurs at t=1s and the islanding detected at t=1.384s so the time detection is Δt=384ms , the islanding detected with the voltage. The THD is equal to 3.9%.

ISSN 1737-9334

d. Slip Mode Frequency Shift (SMS)

The specification of this method is the islanding detection can be with voltage and with the frequency, figure 16 and 17 testify that ,from these two figures we can notice that islanding detected with voltage before islanding detected by frequency but the difference is very small in such a way that we can neglect this difference , also the detection time is Δt=185 ms, the THD which is presented in figure 18 is equal to 2.47% so the most efficient method for micro-inverter is SMS method.

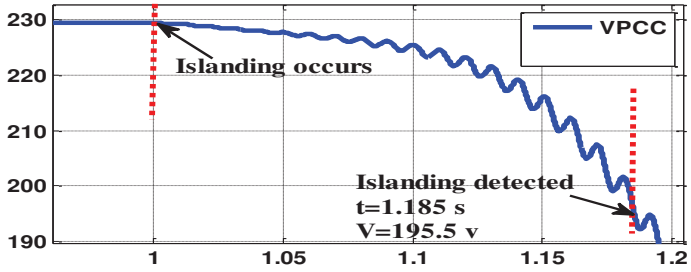


Fig.16: Voltage at PCC change after disconnection from the grid – Method : SMS

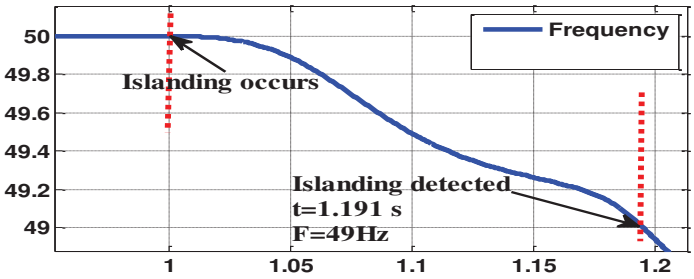


Fig.17: Frequency change after disconnection from the grid – Method: SMS

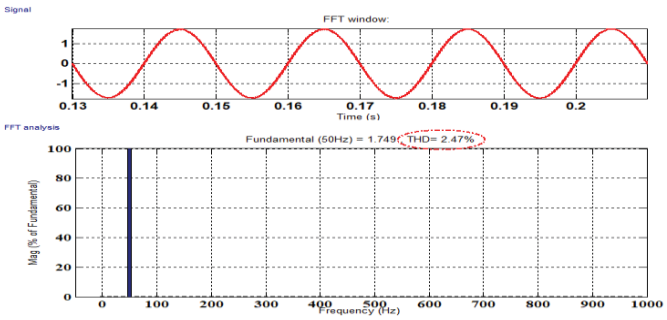


Fig.18: The THD of the output current of the inverter – Method: SMS

V. CONCLUSION

This paper has given a short overview of three active methods used to detect islanding operation for PV inverters and it should show a sense of detection of islanding operation and a simulation principle for these methods.
For better comparison of shown methods it would better to do more simulations for different parameters (for example: load power etc.) and to compare detection time for those parameters. This paper has presented three active methods for islanding detection for micro-inverter 300w, from this work it can be concluded that the SMS method has the better performance for low power inverter with her time detection , his non detective zone, also we can visualize the importance of this method with her resistivity to disturbance with the low THD .

ACKNOWLEDGMENT

This work was supported by the Tunisian Ministry of High Education and Research under Grant LSE-ENITLR 11ES15.

This project is carried out under the MOBIDOC scheme, funded by the EU through the EMORI program and managed by the ANPR.

REFERENCES

[1] IEA (International Energy Agency) Report, ,www.iea.org ,France 2016.
[2] IEEE Standard for Interconnecting Distributed Resources with Electric Power Systems, IEEE Std. 1547-2003, Jul. 2003.
[3] IEEE 929-2000 Standard, "IEEE Recommended Practice for Utility Interface of Photovoltaic (PV) Systems", April 2000
[4] IEEE 1547 Standard, "Interconnecting Distributed Resources with Electric Power Systems", July 2003.
[5] UL1741 Standard, "Inverters, Converters, Controllers and Interconnection System Equipment for Use With Distributed Energy Resources", January 2010.
[6] Ghanbari, T.; Farjah, E.; Naseri, F. ," Power quality improvement of radial feeders using an efficient method. Electr. Power Syst. Res. 163, 140–153,2018.
[7] Canbing Li, Chi Cao, Yijia Cao, Yonghong Kuang, Long Zeng, Baling Fang, "A review of islanding detection methods for microgrid," Renewable and Sustainable Energy Reviews, vol. 35, pp.211-220, 2014
[8] W. Xu, G. Zhang, Li, W. Wang, G. Wang, and J. Kliber, "A power line signaling based technique for anti-islanding protection of distributed generators—Part I: Scheme and analysis," IEEE Trans. Power Del., vol. 22, no. 3, pp. 1758-1766, Jul. 2007.
[9] Andrew J. Roscoe, Graeme M. Burt, Chris G. Bright, "Avoiding the Non-Detection Zone of Passive Loss-of-Mains (Islanding) Relays for Synchronous Generation by Using Low Bandwidth Control Loops and Controlled Reactive Power Mismatches," IEEE Trans. Smart Grid., vol. 5,pp.602-611, 2014
[10] P. Mahat, Z. Chen, B. Bak-jensen, "Review of Island Detection Methods for Distributed Generation", in Proc. of Electric Utility Deregulation and Restructuring and Power Technologies Conference, pp. 2743-2748, 2008.
[11] M. Ropp, D. Larson, S. Meendering, D. MacMahon, J. Ginn, J. Stevens, W. Bower, S. Gonzalez, K. Fennell, L. Brusseau, "Discussion of a Power Line Carrier Communications-based Anti-islanding Scheme using a Commercial Automatic Meter Reading System", in Proc. of Photovoltaic Energy Conversion Conference, pp. 2351-2354, 2006
[12] Z.Ye, A.Kolwalkar, Y.Zhang, P.Du, R. Walling, "Evaluation of Anti-Islanding Schemes Based on Nondetection Zone Concept". IEEE Trans. Power Electron, 19, 1171–1176, 2004
[13] IEA International Energy Agency, Evaluation of Islanding Detection Methods for Photovoltaic Utility Interactive Power systems, in Task V Report IEA PVPS T509: 2002.

ISSN 1737-9334

Optimal Control Through The Minimisation of the Switching Hyperplane

Murat ÜÇÜNCÜ

Electrical-Electronics Department of Engineering Faculty, Başkent University, Bağlıca Kampüsü, Dumlupınar Blv. 20. Km, 06810, Etimesgut /Ankara/Turkey

mucuncu@baskent.edu.tr

***Abstract*—**In an Optimal Control problem, the selection of the matrices in the index of performance is an important problem. The problem becomes more complex as the dimension of the system increases. In most cases, it is not easy to assure a certain dynamic behavior for the evolution of the inputs and the states of the controlled plant. As it is very well known, this type of difficulty can be avoided if the desired indices of performance are specified by the performance of an idealized system called the reference model. In this paper a better approach is presented by the application of Variable Structure System (VSS) theory' results to Continuous and Discrete-Time Optimal Control Systems. It is shown that the application of VSS theory' results in an Optimal Control regulator problem enables the system states to be forced in a region where the system states are insensitive to plant parameter variations and external disturbances while at the same time the states are regulated in an optimal fashion. The design method also provides the flexibility to adjust the system dynamic by adjusting the switching - hyperplane parameters, which are elements in the performance index. Therefore, the selection of the matrix Q is totally replaced by the selection of the switching hyperplane matrix by which it is easy to establish the dynamic behavior of the resulting new system dynamic. Furthermore, the theory is also extended to the Discrete and Continuous Time Stochastic Optimal Control Problems.

***Keywords*—Variable Structure Control, Optimal Control, Continuous Time Control, Discrete Time Control, Stochastic Control**.

I. INTRODUCTION

It is very well known that the application of Variable Structure Systems (VSS) with Sliding modes control results in a new system dynamic which is insensitive to parameter variations and external disturbances. The theory is summarized in the following paragraphs [1],[2],[3]–[4].

Consider a general type of system given below;

$$\dot{x}(t) = A\,x(t) + B\,u(t) \tag{1}$$

$$y(t) = D\,x(t) \tag{2}$$

Suppose that a control input is found such that the system has a sliding regime simultaneously on m switching hyperplanes $s_1(t) = s_2(t) = \ldots\ldots\ldots = s_m(t) = 0$;

where;

$$s_j(t) = \sum_{i=1}^{n} c_{ji}\, x_i(t) = 0 \tag{3}$$

$j = 1, 2, \ldots\ldots\ldots m$

$i = 1, 2, \ldots\ldots\ldots n$

where

$s_{jn} = 1$

In vector notation, this is simply

$$s(t) = C\,x(t) = 0 \tag{4}$$

When there exist a sliding regime simultaneously on all m switching hyperplanes,

$$\frac{ds}{dt} = C\left[A\,x(t) + B\,u(t) = 0\right] \tag{5}$$

Then, the equivalent control is determined as $u_{eq}(t) = -(CB)^{-1}\,C\,A\,x(t)$. In general, the so called equivalent control, $u_{eq}(t)$, is not the actual control applied to the plant. It is only instrumental in finding the sliding mode equations. If this control is substituted into the original system equations supposing that $(CB)^{-1}$ exists, the following VSS sliding mode equations are found.

$$\dot{x}(t) = \left[I_n - B\,(CB)^{-1}\,C\right]A\,x(t) \tag{6}$$

The above equations appear to be of order (nx1), however this is not the case, because due to the sliding $s(t) = C\,x(t) = 0$ is achieved, and therefore m of the state variables in Eq.(3) can be expressed in terms of the remaining (n-m) state variables. It can easily be seen that by adjusting the parameters of switching hyper-plane matrix C, the dynamic of the system with sliding mode control can be adjusted by the designer very easily. In this paper, this idea of forcing the states to move on a set of switching planes is tried to be applied to an optimal control problem where the switching plane is minimized which in turn results in regulating the states as desired by adjusting the parameters of the switching hyperplane matrix, C.

It is stated that $u_{eq}(t)$ is not the actual control applied to the plant and is only instrumental in finding the sliding mode equations. The actual control applied to the system is the so called sliding mode control which is in the following form [1],[2]–[3].

$$u_j(t) = -\sum_{i=1}^{n} (\alpha_j^i\,|x_i(t)| + \delta_j)\,\mathrm{sgn}(s_j) \tag{7}$$

where

ISSN 1737-9334

$$u_j^+(x(t)) \qquad s_j(t) \geq 0 \tag{8}$$

$u_j(t) = \{$

$$u_j^-(x(t)) \qquad s_j(t) \leq 0$$

The control input parameters are evaluated by taking the following conditions into consideration.

$$\dot{s}(t)\,s(t) \prec 0 \tag{9}$$

It is easily seen that if the above inequality is satisfied, then the states will hit the switching hyperplane from any initial conditions and will chatter around it. As a result, we say that $s(t)=0$ is achieved. The VSS theory has also been extended to Discrete Time Systems [5]–[6]. This is also summarized in the following paragraphs;

A single input-single output discrete time system is described by the equations given below.

$$x(k+1) = A\,x(k) + b\,u(k) + d\,f(k) \tag{10}$$

where $u(k)$ is the control input and $f(k)$ is the disturbance added into the system with

$b = (0\ 0\ 0 \ldots\ldots\ldots\ldots b_n)$, $d = (0\ 0\ 0 \ldots\ldots\ldots\ldots d_n)$, A is nxn matrix and $x(k)$ is nx1 state vector.

Suppose that a control input is found such that the states move on the switching plane

$$s(k) = \sum_{i=1}^{n} c_i\,x_i(k) = 0 \tag{11}$$

where c_i : are constants for $i = 1,2,\ldots\ldots n-1$ and . From Eq. (11), we see that $x_n(k) = -\sum_{i=1}^{n-1} c_i\,x_i(k)$

Substituting Eq.(11) into Eq(10), the following new reduced order system equations are obtained as below,

$$x_i(k+1) = \sum_{j=1}^{n-1} (a_{ij} - c_j a_{in})\,x_j(k) \tag{12}$$

$$i = 1,2,\ldots\ldots n-1$$

$$j = 1,2,\ldots\ldots n-1$$

As easily seen, the above equations are insensitive to external disturbances. Since the system is a discrete time system, the reduced order system is stable if the absolute values of the eigenvalues of the new system matrix are less than one. This is simply achieved by adjusting the coefficients of the switching plane.

In a plant of general type, the sliding mode equations can be obtained from the so called equivalent control method. The actual control applied to the system can be found as described in the sequel.

When [1],[2]–[3] is achieved, it has to be maintained. This is possible with a control input which is the solution of

$$s_i(k+1) - s_i(k) = 0 \text{ after having achieved } s_i(k) = 0 \tag{13}$$

If a control is found such that,

$$[s_i(k+1) - s_i(k)]\,s_i(k) \prec 0 \text{ and } |s_i(k+1)| \prec |s_i(k)| \tag{14}$$

which assures both sliding motion and convergence onto the i th hyperplane which can be decomposed into two inequalities [6].

$$[s_i(k+1) - s_i(k)]\,sign(s_i(k)) \prec 0$$

$$[s_i(k+1) + s_i(k)]\,sign(s_i(k)) \leq 0 \tag{15}$$

are satisfied , then the states will hit the switching hyperplane from any initial conditions and will chatter around it. As a result we say that $s(k)=0$ is achieved.

A control input of the following form

$$u(k) = \{ \quad \begin{matrix} u^+(x(k)) & s_i(k) \succ 0 \\ u^-(x(k)) & s_i(k) \leq 0 \end{matrix} \tag{16}$$

is evaluated by taking the above conditions in Eq.16 into consideration.

In this paper, this idea of forcing the states to move on a set of switching planes is tried to be also applied to a discrete time optimal control problem where the switching plane is minimized which in turn results in regulating the states as desired by adjusting the parameters of the switching hyperplane matrix , C .

II. DISCRETE TIME LINEAR REGULATOR PROBLEM

Consider the plant

$$x(k+1) = A\,x(k) + B\,u(k) \tag{17}$$

with a switching hyperplane

$$s(k) = C\,x(k) \tag{18}$$

where $s(k)$ s is (mx1) vector and C is (mxn) matrix.

Define the following performance index so that the switching hyperplane is minimized in an optimal fashion.

$$J_N = \sum_{i=1}^{N} s^T(i)\,s(i) + u^T(i-1)\,R(i-1)\,u(i-1)) \tag{19}$$

We begin by defining V_N to be the minimum value of the performance measure J_N as follows ;

$$V_N = \min_{u(0)u(1)\ldots u(N-1)} \sum_{i=1}^{N} \left[s^T(i)\,s(i) + u^T(i-1)\,R(i-1)\,u(i-1)\right] \tag{20}$$

Using the principle of optimality, we proceed by starting with the last stage of control in our problem.

$$V_1 = \min_{u(N-1)} \left(s^T(N)\,s(N) + u^T(N-1)\,R(N-1)\,u(N-1)\right) \tag{21}$$

where

$$s(N) = C\,x(N)$$

$$x(N) = A\,x(N-1) + Bu(N-1) \tag{22}$$

Upon substitution in Eq.22, we obtain

$$s(N) = C\{A\,x(N-1) + Bu(N-1)\}$$

Then, V_1 becomes

ISSN 1737-9334

$$V_1 = \min_{u(N-1)} \left\{ \begin{matrix} \left(A\,x(N-1)+Bu(N-1)\right)^T C^T C \\ \left(A\,x(N-1)+Bu(N-1)\right)+u^T(N-1) \\ R(N-1)u(N-1) \end{matrix} \right\}$$

If we drop the time argument for simplicity,

$$V_1 = \min_{u(N-1)} (x^T A^T C^T CAx + x^T A^T C^T CBu + u^T B^T C^T CAx + u^T B^T CC^T Bu + u^T Ru)$$

Denote $Q = C^T C$, it is easily seen that Q is automatically a positive semidefinite symmetric matrix,

$$V_1 = \min_{u(N-1)} (x^T A^T QAx + x^T A^T QBu + u^T B^T QAx + u^T (B^T QB + R)u \quad (23)$$

Since Q is symmetric,

$$(u^T B^T QAx)^T = x^T A^T Q^T Bu = x^T A^T QBu$$

Then, the third term in Eq.23 is the transpose of the second term. Since both are scalars, the two terms are equal. Therefore, we write

$$V_1 = \min_{u(N-1)} (x^T A^T QAx + 2x^T A^T QBu + u^T (B^T QB + R)u) \quad (24)$$

We obtain the minimum in Eq.24 by setting the gradient of the terms with respect to u equal to zero. Then, we have

$$2x^T A^T QB + 2u^T (B^T QB + R) = 0$$

Solving for u we see that

$$u(N-1) = -(B^T QB + R)^{-1} B^T QA\,x(N-1) = 0$$

As it is seen, if R is selected as a positive definite matrix, the resulting control law is physically realizable and additionally is linear and involves feedback of the current state.

We define

$$L(N-1) = -(B^T QB + R)^{-1} B^T Q$$

Then,

$$u(N-1) = L(N-1)\,x(N-1)$$

As the reader will readily recall, in the discrete time optimal regulator problem, the following performance measure is selected [7].

$$J_N = \sum_{i=1}^{N} x^T(i)\,G(i)\,x(i) + u^T(i-1)\,R(i-1)\,u(i-1)) \quad (25)$$

For the plant in Eq.17, if we evaluate V_1 for the above J_N , it becomes

$$V_1 = \min_{u(N-1)} (x^T A^T GAx + 2x^T A^T GBu + u^T (B^T QB + R)u)$$

This is the same as Eq.24 with the exception that G is replaced by Q which is the product of the switching hyperplane matrix by its transpose. i.e CC^T. The design approach is different in our case. Rather than selecting G being at least a positive semidefinite matrix and selecting a performance index as in Eq.25 which is to regulate the states, we attempt to minimize the switching plane resulting in a new desired system dynamic and force the system states in a region where the system is less sensitive to plant parameter variations and external disturbances. If we continue to derive the performance measure for the N-stages, we get the following equations.

$$u(k) = L(k)\,x(k)$$

$$L(k) = -\left(B^T W(k+1)B + R\right)B^T W(k+1)A$$

$$W(k) = A^T W(k+1)A + A^T W(k+1)BL(k) + Q(k)$$ for k=N-1, N-2, ... 0.

where

$$W(N) = Q(N)$$

$B^T W(k+1)B + R$ is required to be positive definite for all k.

III. CONTINUOUS TIME LINEAR REGULATOR PROBLEM

In this section, we shall consider one class of optimal control problem, the linear regulator problem, in which we shall employ the VSS theory by selecting a performance measure by which the switching plane is tried to be minimized, which in turn results in state regulation in a region where there is insensitivity to parameter variations and external disturbances. Namely, the same idea followed in the previous section is followed here and applied to a continuous time linear regulator problem. The plant to be considered is described by the continuous-time linear state equations given below ;

$$\dot{x}(t) = A\,x(t) + B\,u(t) \quad (26)$$

$$y(t) = D\,x(t)$$

The performance measure to be minimized is selected as follows ;

$$J = \frac{1}{2}s^T(t_f)\,s(t_f) + \frac{1}{2}\int_0^t \left[s^T(t)s(t) + u^T(t)\,R(t)\,u(t)\right]dt \quad (27)$$

The physical interpretation of this performance measure is that the switching plane is reached and minimized without an excessive expenditure of control effort, which in turn results in regulating the states in a region where the system is insensitive to parameter variations and external disturbances. Besides, the minimization of the switching hyperplane is achieved without chattering. One way to solve the above minimization problem is to select the Hamiltonian equations as follows;

$$H\left[x(t),u(t),p(t),t\right] = \frac{1}{2}s^T(t)s(t) + \frac{1}{2}u^T(t)\,R(t)\,u(t) + p^T(t)\,A(t)x(t) + p^T(t)\,B(t)\,x(t) \quad (28)$$

where

$$s(t) = C^T(t)\,x(t)$$

Then, the optimal control which minimizes the above Hamiltonian can be found by using the parallel approach defined in [8] as follows ;

$$u(t) = -R^{-1}B^T K(t)x(t) \quad (29)$$

$$\dot{K}(t) = -K(t)A(t) - A^T K(t) - C(t)C^T(t) +$$

ISSN 1737-9334

$$K(t)B(t)R^{-1}(t)B^T(t)K(t) \quad (30)$$

where $K(t)$ is the gain matrix and can be evaluated as described in [8]. This is a slight modification to the original linear regulator problem and the control input evaluated by this approach facilitates the robust properties of VSS.

IV. STOCHASTIC CONTINUOUS TIME LINEAR REGULATOR PROBLEM

In the previous cases, it is assumed that all the states are available. This assumption may not be valid in practical applications. Furthermore, the processes may be stochastic. In such cases, one method is to estimate the states and use the estimated values of the states to evaluate the switching plane value, which is in fact the estimated value of the switching plane. To apply this idea, we extend the idea to the minimization of the estimate of the switching plane as described in the following paragraphs.

We consider the following system ,

$$\dot{x}(t) = A\,x(t) + B\,u(t) + G(t)w(t) \quad (31)$$
$$y(t) = H\,x(t) + v(t)$$

The stochastic processes $\{ w(t), t \geq t_0 \}$ and $\{ v(t), t \geq t_0 \}$ are assumed to be zero mean Gaussian white noises with

$$E\left[w(t)w^T(\tau)\right] = Q(t)\,\delta(t-\tau)$$
$$E\left[w(t)w^T(\tau)\right] = R(t)\,\delta(t-\tau)$$

and

$$E\left[w(t)v^T(\tau)\right] = 0$$

for all $t, \tau \geq t_0$ where all the terms have been defined previously .

If a sliding mode control which satisfies

$$\dot{s}(t)\,s(t) \prec 0$$

where $s(t) = C\,x(t)$ as described in Eq. 7 is applied to the above system ,the system after a certain time moves on the switching plane $s(t) = 0$

The equivalent control can be found as follows,

$$u_{eq}(t) = -(CB)^{-1}\,C\left[A\,x(t) + C\,G\,w(t)\right]$$

Upon substitution into the original system equations, Eq.31, the new dynamic is obtained as follows,

$$\dot{x}(t) = \left[I_n - B(CB)^{-1}C\right]\left[A\,x(t) + C\,G\,w(t)\right]$$

For the total disturbance rejection, the switching plane matrix must be chosen such that

$$\left[I_n - B(CB)^{-1}C\right]CG = 0$$

This gives

$$\dot{x}(t) = \left[I_n - B(CB)^{-1}C\right]Ax(t) \quad (32)$$

Then, the desired motion is achieved by adjusting the coefficients of C.

Rather than selecting a sling mode control , the switching plane may try to be minimized , which in fact enables the regulation of s(t) and achieves s(t) = 0, when in turn , the system states move on the switching plane resulting in a new and desired dynamic according to the selection of the switching hyperplane. To achieve this purpose, the following performance measure to be minimized is selected as follows ;

$$J = E\left\{ s^T(t_f)\,s(t_f) + \int_0^t \left[s^T(t)s(t) + u^T(t)\Gamma(t)u(t) \right] dt \right\} \quad (33$$

E {s(t)} = E {s(t/t)} : Expected value of the switching hyperplane given all the measurements up to t.

If a control input is found to minimize the expected value of the above switching plane, then the expected values of the states will be forced to move in a region where we have the robust properties of the VSS, while at the same time, the estimated states are regulated in an optimal fashion. Thus, If $s(t) = C\,x(t)$ is substituted in the above performance index for s(t), the following equivalent performance index is obtained.

$$J = E\left[x^T(t_f)C^T C\,x(t_f)\right] + E\left[\int_0^t x^T(t)C^T C\,x(t)\,dt\right]$$
$$+ E\left[\int_0^t u^T(t)\Gamma(t)u(t)\,dt\right] \quad (34)$$

If the similar steps described in [7] are followed, the optimal control for the continuous stochastic linear regulator problem can be found to be characterized by the set of following relations,

$$u(t) = L(t)\,x(t/t)$$
$$L(t) = -\Gamma^{-1}B^T(t)W(t)$$
$$\dot{W}(t) = -A^T W - WA + WB\Gamma^{-1}B^T(t)W(t) - C^T(t)C(t)$$
$$\dot{x}(t/t) = Ax(t/t) + K(t)\left[z(t) - H(t)x(t/t)\right] + Bu(t)$$

It is interesting to note that in the classical VSS the measurement of the output is not taken into consideration and it is assumed that all the system states are assumed to be available. In the approach explained above, it is clear that the output measurement is taken into consideration.

V. STOCHASTIC DISCRETE TIME LINEAR REGULATOR PROBLEM

We can extend the same argument to the discrete time stochastic linear regulator problem by following the parallel approach studied in the previous paragraphs.

We again select the following performance for the similar reasons discussed in the previous paragraphs.

$$J_N = E\left\{ \sum_{i=1}^{N} s^T(i)\,s(i) + u^T(i-1)\,R(i-1)\,u(i-1) \right\}$$

If the similar steps described in [7] are followed for a plant in Eq.17, the optimal control for the discrete time stochastic linear regulator problem can be found to be characterized by the following set of relations:

ISSN 1737-9334

$$u(k) = L(k)\,x(k/k)$$

$$L(k) = -\left[B^T W(k+1)B + R\right]^{-1} B^T W(k+1)A$$

$$L(k) = -\left[B^T W(k+1)B + R\right]^{-1} B^T W(k+1)A$$

$$W(k) = A^T W(k+1)A + A^T W(k+1)BL(k) + Q(k)$$ for k=N-1, N-2, ... 0.

where

$$W(N) = Q(N)$$

$B^T W(k+1)B + R$ is required to be positive definite for all k.

$$\hat{x}(k/k) = \hat{x}(k/k-1) + K(k)\left[z(k) - H(k)\hat{x}(k/k-1)\right]$$

$$\hat{x}(k/k-1) = A\hat{x}(k-1/k-1) + Bu(k-1)$$

where K(k) is the Kalman gain matrix described in [7]

V.CONCLUSION

The most favorable aspect of variable structure controller is that the new resulting system dynamic is insensitive to plant parameter variations and external disturbances. In this paper, this robust property is extended to the conventional state regulation problem. It is shown that the states of the system can be regulated in an optimal fashion by minimizing the switching hyperplane. This approach facilitates the design of a more robust optimal control regulator by combining and facilitating the use of VSS theory. In practical applications, the system states are random. Therefore, the estimated states can only be used in the evaluation of the switching hyperplane. Therefore, the theory is also extended to cover the stochastic cases.

REFERENCES

[1] U.Itkis, Control Systems of Variable Structure., Newyork : Wiley, 1976.

[2] V.I. Utkin, Sliding Modes and Their Application in Variable Structure Systems. Moscow: Mir, 1978.

[3] V.I. Utkin, “Variable Structure Systems with Sliding Modes, IEEE Trans. Automat. Contr., Vol.AC-22, pp.212-22, 1977.

[4] Kar-Keung D. Young, “Controller Design for a manipulator using theory of variable structure systems”, IEEE Trans. Systems, Man and Cybernetics, vol. SMC-8, pp.101-109, 1978.

[5] Murat Üçüncü and Yorgo Istefanopulos, “Discrete Time Variable Structure Control”, Measurement and Control Conference, 1985, İstanbul-Turkey.

[6] Sami Sarpkaya, Bosphorus University, EE Department, Ph.D. Thesis, 1990, Istanbul-Turkey.

[7] Meditch, “Stochastic Optimal Linear Estimation and Control”, McGraw-Hill Series.

[8] Donald E.Kirk, Optimal Control Theory, Prentice hall Network Series.

[9] Arthur Gelb, “Applied Optimal Estimation”, ISBN 0-262-20027-9.

ISSN 1737-9334

Sliding Mode Control Design Based on the LMI Approach for GMAW Welding Process

Sami Kahla[1], Amar Boutaghane[1], Said Dehimi[1],Ahmed Kellai, Noureddine Hamouda[1],Babes Badreddine[1], Rachid Amraoui[1]

[1]*Research Center in Industrial Technologies, CRTI, P.O. Box 64, 16014 Cheraga, Algiers, Algeria.*

s.kahla@crti.dz, a.boutaghane@crti.dz,s.dehimi@crti.dz, a.kellai@crti.dz, n.hamouda@crti.dz, b.babes@crti.dz,r.amraoui@crti.dz

***Abstract*— Gas metal arc welding (GMAW) plays the great importance in the welding industry. This paper explores the design of a sliding matrix controller based on linear matrix inequality (LMI) for an uncertain gas metal arc welding system (GMAW). The results obtained can ensure a better dynamic behavior of the GMAW process. The local stability of the closed-loop system is analyzed with the presence of uncertainties in the linearized model of the process as well as the control parameters. Simulation and experimental results reveal that the proposed controller enjoys adequate performance and stability.**

***Keywords*— GMAW, MIMO system, Sliding Mode Control, Linear Matrix Inequality (LMI).**

I. INTRODUCTION

Welding is an assembly technique used in the most industrial sectors. It is a multidisciplinary technique mobilizing a large number of phenomena [1]. Among many types of welding processes, gas arc welding (GMAW) is very attractive because of its high speed and its both manual and automatic welding modes for a wide range of ferrous and non-ferrous metals [2].

In the most of automatic welding systems, voltage and arc current are the main variables controlled. In fact, these two variables affect many characteristics of the weld, such as the geometry of the weld [3], the metallurgical characteristics of the weld [4], mode of transfer of the molten droplets [5], weld stability [6], weld defects and weld quality [7]. Until now, different control approaches are implemented to control these two important variables. In reference [2], the authors designed two PI controllers to control the process. Since the process behaves non-linearly within its operational range and there is strong interaction between its variables, it would be difficult to adjust the PI parameters to achieve the desired performance throughout the entire operating range of the process.

In order to evaluate and ensure the weld quality in the GMAW process, arc length and current are important controlled variables. These two variables of welding is determined by several characteristics such as the transfer mode of melting droplets and the weld geometry [8], in this reason the control of the GMAW process can be separated into weld pool control and arc control [9]. Previous studies for the GMAW process have been implemented to control the arc length and current; in [10, 11] a controller design has been already combined with feedback linearisation technique where non linearities are cancelled and the linearised strategy is managed by (PI) controller. A similar strategy with sliding mode control has been applied to ensure robustness in [12].

Recently, sliding mode control is a well accepted technique for the systematic design of robust controllers for complex non-linear dynamic systems operating under conditions of uncertainty. Thomsen proposed a control system for manual pulsed gas metal arc welding. The particular arc length controller is dependent on a non-linear SISO model of the arc length process and uses feedback linearization approach [13].

Khatamianfar et al are proposed a novel application of sliding control in the manual gas metal arc welding process; Arc length is controlled successfully by the robust control system with combined the feedback linearization technique and sliding mode control [14]. Golob. M [15] combined a full-bridge inverter circuit together with the GMAW model.

The goal of this work is to maintain a desired current and arc length by designing sliding mode controllers based on the linear matrix inequality (LMI) approach for the linear model of the GMAW system. This model is suitable for manual welding where extreme robustness against parameter variations and external disturbances is of great importance. The desired arc length must be maintained despite unexpected changes in work piece distance, considered an external disturbance.

This paper is organized as follows: First in section 2, the mathematical modeling of a GMAW process is presented and, then in section 3, the control objective is discussed. Subsequently, a sliding mode control based on the linear matrix inequality is designed. Applications of the sliding mode control based on the LMI to the GMAW process and simulation results are given in the section 4.Finaly, the conclusions are drawn.

II. MODELING OF THE GMAW PROCESS

The schematic diagram of the GMAW system is illustrated in the Figure 1. The power source consists of a constant voltage source connected to the electrode and the work piece. The wire speed, S, travel speed of the torch, R, open circuit voltage Voc, and contact tip to work piece distance, CT, are adjusted to get the desired weld. The model used in this work is the fourth-generation of the derivative equation that originated at the Idaho National Engineering and Environmental Laboratory (INEEL) [16, 17].

The main parts of the model will be presented as follows. Basically, the important aspects with respect to control are the

ISSN 1737-9334

electrical circuit, the drop dynamics, the drop detachment criteria and the melting rate.

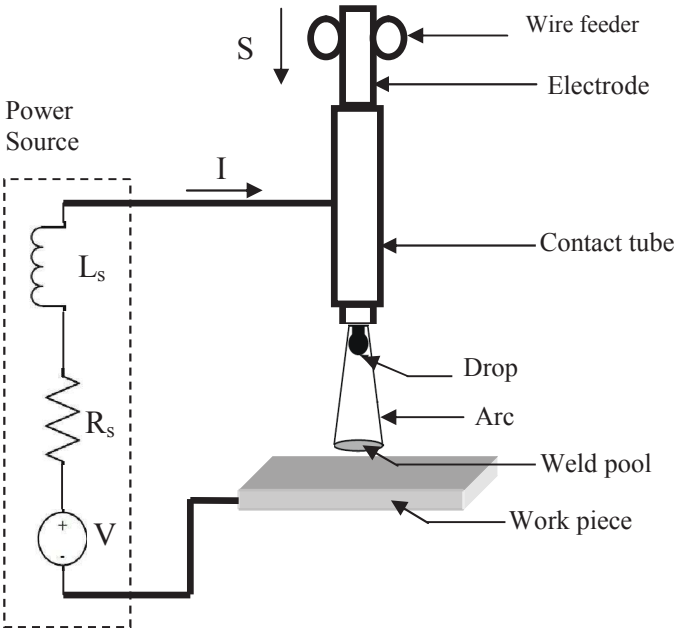


Figure 1.Schematic diagram of GMAW system.

The current is obtained from a simple electrical circuit as [18]:

$$\dot{I} = \frac{V_{oc} - R_L I - V_{arc} - R_s I}{L_s} \quad (1)$$

$$V_{arc} = V_o + R_a I + E_a(CT - l_s) \quad (2)$$

$$R_L = \rho\left[l_s + \frac{1}{2}(r_d + x_d)\right] \quad (3)$$

The pendant drop attached to the tip of the electrode can be modelled as a mass-spring-damper system as described by [19] and Wu et al given by the following equation:

$$m_d\ddot{x}_d = F_T - b_d\dot{x}_d - k_d x_d \quad (4)$$

Where F_T is the total external force affecting the droplets is as follows:

$$F_T = F_g + F_{em} + F_d + F_m \quad (5)$$

Where F_g, F_{em}, F_d and F_m are the gravity force, electromagnetic force, plasma drag force, and momentum flux force, respectively. The gravity force F_g is defined as

$$F_g = m_d \quad (6)$$

Whereg is the gravity and m_d is the mass of the droplet, which can be computed in terms of the droplet radius r_d . It is assumed that the droplets have a spherical shape.

$$m_d = \frac{4}{3}\pi\rho_e r_d^3 \quad (7)$$

ρ_e describes the density of the liquid electrode material. Considering uniform distributed current and a spherical droplet with a radius larger than the solid electrode, the electromagnetic force F_{em} can be written as

$$F_{em} = \frac{\mu_0 I^2}{4\pi}\left[\ln\left(\frac{r_d \sin\theta}{r_e}\right) - \frac{1}{4} - \frac{1}{1-\cos\theta} + \frac{2}{(1-\cos\theta)^2}\ln\left(\frac{2}{1+\cos\theta}\right)\right] \quad (8)$$

Where μ_0 is the permeability of free space, I is the welding current, θ is the angle of the arc-covered area, and r_e is the radius of the electrode . The plasma drag force F_d is determined as follows.

$$F_d = 0.5\left(c_d A_d \rho_p v_p^2\right) \quad (9)$$

Where c_d is the drag coefficient, ρ_p is the density of the plasma, v_p is the shielding gas velocity, and A_d is the area of droplet hit by the shielding gas

$$A_d = \pi(r_d^2 - r_e^2) \quad (10)$$

The momentum flux F_m is determined as follows [20].

$$F_m = \frac{\mu_0}{4\pi}\left(\left(\frac{I}{\sigma}\right)^2 - I_2\right) \quad (11)$$

Whereσ is defined as r_d/r_e and I_2 is determined as

$$I_2 = \int J_0 dA_d \quad (12)$$

$J_0$is the uniform current density on the arc covered area of the drop surface .

The melting rate M_R is expressed by the following equation:

$$M_R = C_2 I_a^2 \rho l_s + C_1 \quad (13)$$

The stick-out evolution is controlled by

$$\frac{dl_s}{dt} = S - \frac{M_R}{\pi r_e^2} \quad (14)$$

The state-space representations of the resulting equation are given in the following equations. Firstly, the state variables are defined as:

$x_1 = x_d$ Droplet displacement (m).
$x_2 = \dot{x}_d$ Droplet velocity (m/sec).
$x_3 = m_d$ Droplet mass (kg).
$x_4 = l_s$ Stick-out (m).
$x_5 = I$ Current (A).

Where x_d is the distance of the center of the mass of the droplet above the work piece.

Then the nonlinear state equations can be written as follows:

$$\begin{aligned}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= \frac{-kx_1 - Bx_2 + F_T}{x_3} \\
\dot{x}_3 &= M_R\rho_e \\
\dot{x}_4 &= u_1 - \frac{M_R}{\pi r_e^2} \\
\dot{x}_5 &= \frac{u_2 - (R_a + R_s + R_L)x_5 + V_0 - E_a(CT - x_4)}{L_s}
\end{aligned} \quad (15)$$

Where k and B are the spring constant and damping coefficient of the droplet.R_a and R_s are the arc resistance and source resistance, respectively. $\rho_e$is the electrode density, V_0 is the arc voltage constant, E_a is the arc length factor and L_s is the source inductance.

The output equations are given by the following equations:

$$\begin{aligned}
y_1 &= l_a \\
y_2 &= I_a
\end{aligned} \quad (16)$$

And the control variables are

$u_1 = S$ Wire feed speed (m/sec),
$u_2 = V_{oc}$ Open-circuit voltage (V).

ISSN 1737-9334

States of the system must be reset after each detachment of a drop, which means:

$$F_T > F_s \quad (17)$$

Or

$$r_d > \frac{\pi(r_d+r_e)}{1.25(\frac{x+r_d}{r_d})(1+\frac{\mu_0 I^2}{2\pi^2\gamma(r_d+r_e)})^{\frac{1}{2}}} \quad (18)$$

Where:

$$r_d = \left(\frac{3x_5}{4\pi\rho_w}\right)^{\frac{1}{3}} \quad (19)$$

And F_s is the surface tension of the droplet given as:

$$F_s = 2\pi\gamma r_e \quad (20)$$

Where γ is the surface tension of liquid steel [20].

Table. 1 gives the numerical data for the nonlinear model parameters.

By linearizing the equations of the system around an operating point, the representation of the process state is obtained as follows:

$$\begin{cases}\dot{x} = Ax + Bu \\ y = Cx\end{cases} \quad (21)$$

Where

$$x = \begin{bmatrix}\delta l_s \\ \delta i_a\end{bmatrix}, \; u = \begin{bmatrix}\delta s \\ \delta v_{oc}\end{bmatrix}, \text{ et } y = \begin{bmatrix}\delta l_s \\ \delta i_a\end{bmatrix}.$$

As usual, lowercase letters have been used here to represent linearized model variables. δrepresents the variations around the operating point. The matrices of the linearized model are determined as follows:

$$A = \begin{bmatrix} -\frac{c_2\rho}{\pi r_e^2}\bar{i}_a^2 & -\frac{c_1+2c_2\rho\bar{i}_a\bar{l}_s}{\pi r_e^2} \\ L_s^{-1}(E_a - \rho\bar{i}_a) & -L_s^{-1}\left(R_a + R_s + \rho\bar{l}_s\right)\end{bmatrix}$$

$$B = \begin{bmatrix}0 & 0 \\ 0 & L_s^{-1}\end{bmatrix}, \quad \text{et} \quad C = \begin{bmatrix}1 & 0 \\ 0 & 1\end{bmatrix} \quad (22)$$

Superscript $^-$ is used to display the values of the variables at the operating point.

III. Sliding mode controller design based on linear matrix inequality (LMI) for a GMAW welding process

A. *Sliding mode controller design based on linear matrix inequality*

Consider the linear system defined by equation (21)

$$\begin{cases}\dot{x} = Ax + Bu \\ y = Cx\end{cases}$$

where $x \in R^n$, $u \in R^n$, $A \in R^{n\times n}$, $B \in R^{n\times n}$.

The control objective is $x \to x_r$, x_r is the ideal reference. Set the tracking error on $z = x - x_r$,and then

$$\dot{z} = \dot{x} - \dot{x}_r = Ax + Bu - \dot{x}_r \quad (23)$$

The controller is designed as

$$u = Fx + u_r \quad (24)$$

Where F is the state feedback gain, which can be solved by LMI, $u = Fx - Fx_r - B^{-1}Ax_r + B^{-1}\dot{x}_r = Fz - B^{-1}Ax_r + B^{-1}\dot{x}_r$, then

And

$$\dot{z} = Ax + B(Fz - B^{-1}Ax_r + B^{-1}\dot{x}_r) - \dot{x}_r = Az + BFz \quad (25)$$

Table 1.Parameters and variables of the GMAW process

Nomenclature	Symbol	Value (unit)
Source resistance	R_s	$6.8\text{x}10^{-3}(\Omega)$
Source inductance	L_s	$306\text{x}10^{-6}$(H)
Arc resistance	R_a	0.0237 (Ω)
Arc length factor	E_a	400 (V m^{-1})
Contact tip to work piece distance	l_c	0.025 (m)
Arc voltage constant	V_0	15.5 (V)
Permeability of free space	μ_0	$4\pi \times 10^{-7}$ (H m^{-1})
Gravity	g	9.8 (m s^{-2})
Spring constant of drop	k_d	3.5 (kg s^{-2})
Damper constant of drop	b_d	0.8×10^{-3} (kg s^{-1})
Density of the liquid electrode material	ρ_e	7860 (kg m^{-3})
Electrode radius	r_e	0.006 (m)
Shielding gas velocity	v_p	10 (m s^{-1})
Density of the plasma	ρ_p	1.784 (kg m^{-3})
Drag coefficient	c_d	0.44
Melting rate constant 1	c_1	3.3×10^{-10} (m^3 s^{-1} A^{-1})
Melting rate constant 2	c_2	0.78×10^{-10} (m^3 s^{-1} Ω^{-1} A^{-2})
Surface tension of liquid steel	γ	1.3 (N m^{-1})

Theorem. 1: If the following LMI is satisfied

$$\alpha P + A^T P + M^T + PA + M < 0 \quad (26)$$

where $F = (PB)^{-1}M$.

Then, the closed-loop (21) and controller (24) system is exponentially stable.

Proof.1: Choose Lyapunov function as

$$V = z^T Pz$$

With $P = P^T > 0$

Then

$$\dot{x} = Ax + BFx$$

And

$$\begin{aligned}\alpha V + \dot{V} &= \alpha V + (z^T P)'z + z^T P\dot{z} = \alpha V + \dot{z}^T Pz + z^T P\dot{z} \\ &= \alpha V + (Az + BFz)^T Pz + z^T P(Az + BFz) \\ &= \alpha V + z^T A^T Pz + z^T F^T B^T Pz + z^T PAz + z^T PBFz \\ &= \alpha V + z^T(A^T P + F^T B^T P + PA + PBF)z \\ &= \alpha z^T Pz + z^T \Omega z = x^T(\alpha P + \Omega)z\end{aligned}$$

Where $\alpha > 0$, $\Omega = A^T P + F^T B^T P + PA + PBF$.

To ensure that $\alpha V + \dot{V} \le 0$, $\alpha P + \Omega < 0$ is necessary, that is to say:

$$\alpha P + A^T P + F^T B^T P + PA + PBF < 0 \quad (27)$$

Since F et P are unknown, solve Eq. (27), the left part must be linearized. To define $M = PBF$, so we have

$$\alpha P + A^T P + M^T + PA + M < 0 \quad (28)$$

From $\alpha V + \dot{V} \le 0$, we have $V(t) \le V(0)\exp(-\alpha t)$,i.e.,if $t \to \infty, V(t) \to 0$ et $z \to 0$, the system declares convergence to zero exponentially.

B. *Sliding mode controller design based on linear matrix inequality with disturbance*

Consider the linear system defined by

$$\begin{cases}\dot{x} = Ax + Bu + d \\ y = Cx\end{cases} \quad (29)$$

Where $x \in R^n$, $u \in R^n$, $A \in R^{n\times n}$, $B \in R^{n\times n}$, $d \in R^{n\times 1}$ is disturbance.

Set the tracking error on $z = x - x_r$,and then

$$\dot{z} = \dot{x} - \dot{x}_r = Ax + Bu + d - \dot{x}_r \quad (30)$$

The controller is designed as

ISSN 1737-9334

$$u = Fx + u_r + u_s \quad (31)$$

Where F is the state feedback gain, which can be solved by LMI, $u_r = -Fx_r - B^{-1}Ax_r + B^{-1}\dot{x}_r$, $u_s = -B^{-1}(\eta sgn(z))$, $\eta \in R^{n\times 1}$, $\eta_i > \bar{d}_i$, $\eta sgn(z) = [\eta_1 sgnz_1 \quad \cdots \quad \eta_n sgnz_n]^T$.

then

$$u = Fx - Fx_r - B^{-1}Ax_r + B^{-1}\dot{x}_r - B^{-1}(\eta sgn(z))$$

And

$\dot{z} = Ax + B\big(Fz - B^{-1}Ax_r + B^{-1}\dot{x}_r - B^{-1}(\eta sgn(z))\big) + d - \dot{x}_r$.

$= Ax + BFz - Ax_r - \eta sgn(z) + d$.

Theorem. 2 [21]: If the following LMI is satisfied

$$\alpha P + A^T P + M^T + PA + M < 0 \quad (32)$$

where $\alpha > 0$, $F = (PB)^{-1}M$.

Then, the closed-loop (29) and controller (31) system is exponentially stable.

Proof.2: Choose Lyapunov function as

$$V = z^T Pz$$

With $P = diag(p_i)$ is a diagonal matrix and $p_i > 0$.

Then

$$\begin{aligned} \alpha V + \dot{V} &= \alpha V + (z^T P)'z + z^T P\dot{z} = \alpha V + \dot{z}^T Pz + z^T P\dot{z} \\ &= \alpha V + (Az + BFz - \eta sgn(z) + d)^T Pz \\ &\quad + z^T P(Az + BFz - \eta sgn(z) + d) \\ &= \alpha V + z^T A^T Pz + z^T F^T B^T Pz + (-\eta sgn(z) + d)^T Pz \\ &\quad + z^T PAz + z^T PBFz + z^T P(-\eta sgn(z) \\ &\quad + d) \\ &\le \alpha V + z^T(A^T P + F^T B^T P + PA + PBF)z \\ &= \alpha z^T Pz + z^T \Omega z = z^T(\alpha P + \Omega)z \end{aligned}$$

Where $\alpha > 0$, $\Omega = A^T P + F^T B^T P + PA + PBF$.

$(-\eta sgn(z) + d)^T Pz = \sum_{i=1}^{n}(-\eta_i + d_i)p_i|z_i| < 0$,

$z^T P(-\eta sgn(z) + d) = \sum_{i=1}^{n}(-\eta_i + d_i)p_i|z_i| < 0$.

To ensure that $\alpha V + \dot{V} \le 0$, $\alpha P + \Omega < 0$ is necessary, that is to say:

$$\alpha P + A^T P + F^T B^T P + PA + PBF < 0 \quad (33)$$

From $\alpha V + \dot{V} \le 0$, we have $V(t) \le V(0)\exp(-\alpha t)$,i.e.,if $t \to \infty, V(t) \to 0$ et $z \to 0$, the system declares convergence to zero exponentially.

IV. Process Simulation and Results

Design procedures are implemented and simulation studies are conducted using MATLAB. All of the simulation results presented in this paper were obtained by solving the linear process state space model with a sampling time of 50 ms and implementing the controllers with a sampling interval of 0.02 s. The working point in a GMAW process is selected according to the transfer mode of the droplets (which may be short circuits, globular and spray), which itself depends on the material and the thickness of the part. Previously, the model used in this paper represents the mode of pulverized metal transfer mode. Thus, it is necessary to select a working point such that the GMAW process operates in this transfer mode. The selected operating points are:

$$\bar{I}_a = 230\ A, \bar{l}_s = 15mm, \bar{V}_{oc} = 28\ V, \bar{l}_c = 25mm \quad (34)$$

The step operating point variations in (34) with ± 1mm step sizes for arc length and ± 5A for arc current are considered as closed loop system reference inputs. The first two seconds of the simulation results are mainly affected by the initial values of the states, while the effects of the controllers appear mainly afterwards. The matrices of the linearized model are determined as follows:

$$A = \begin{bmatrix} -\dfrac{c_2\rho}{\pi r_e^2}\bar{X}_2^2 & -\dfrac{c_1 + 2c_2\rho\bar{X}_1\bar{X}_2}{\pi r_e^2} \\ L_s^{-1}(E_a - \rho\bar{X}_2) & -L_s^{-1}(R_a + R_s + \rho\bar{X}_1) \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 0 \\ 0 & L_s^{-1} \end{bmatrix}, \quad \text{et} \quad C = \begin{bmatrix} -E_a & R_a \\ 0 & 1 \end{bmatrix}$$

And for the operating points given by equation (34), applying the LMI approach by equation (27), we can have:

$$F = \begin{bmatrix} -498.22 & 301 \\ 301 & -0.11 \end{bmatrix}$$

Figures 2-3 show respectively arc current and arc length obtained at the nominal values of the GMAW parameters. Figure 2 shows that the arc current has been appropriately controlled and follows the corresponding variable in the linear model of the process, both transient and stable. According to Figure 3, the arc length is unbiased and correctly follows the corresponding reference in the linear model of the GMAW process.

To test the robustness of the performances of the proposed scheme, the following perturbation $d(t) = [50\sin t \quad 50\,sin\,t]$ is considered. The simulation results are given in Figures 4-5.

In this case, in the presence of the perturbation, we can also notice that the arc current and the arc length respectively converge to the desired reference. Then, we can conclude that the tracking performance with the controller proposed in this simulation demonstrates the robustness of the controller.

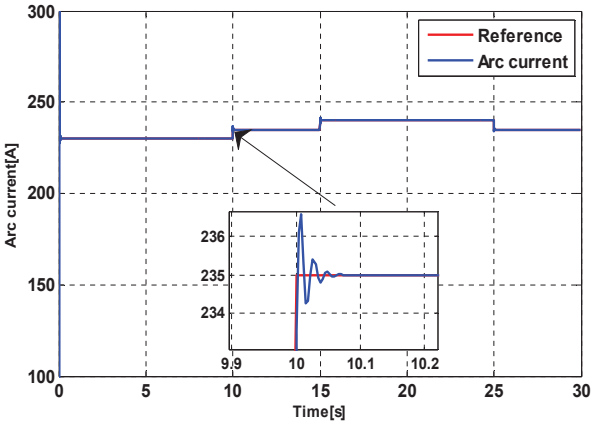


Figure 2.Arc Current based LMI sliding mode controller

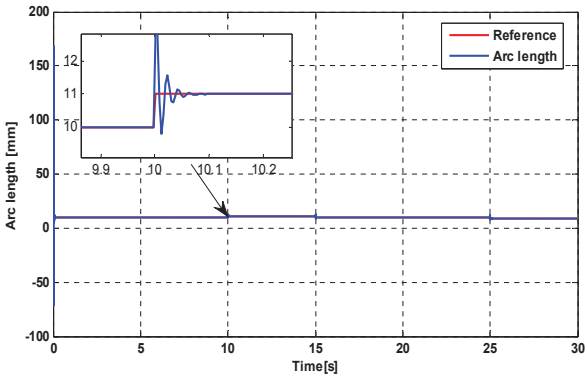


Figure 3.Arc length based LMI sliding mode controller

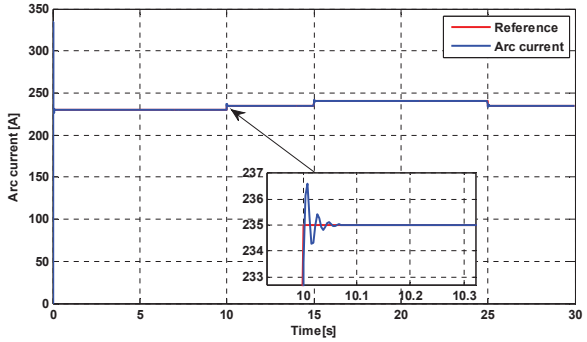


Figure 4. Arc Current based LMI sliding mode controller

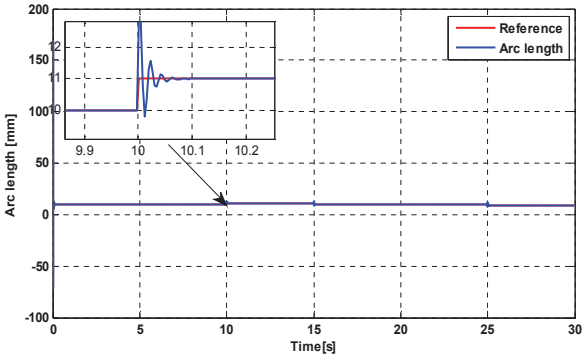


Figure 5. Arc length based LMI sliding mode controller

V. Conclusions

This paper discusses linearized sliding mode control based on the LMI approach for an uncertain GMAW system. The proposed strategy has proven effective for this complex application, showing extremely robust behavior to accurately track arc length and current. The use of the sliding mode controller based on the LMI approach has been studied in detail in this paper. To evaluate and compare simulation results, some well-known measures have been taken into account. In summary, the sliding-mode controller based on the proposed LMI approach performed other controllers in terms of transient response, desired output tracking, and robustness to process parameter uncertainties. In another case where the proposed controller was not the best, it was not too bad compared to the best controller.

References

[1] T. DebRoy: Physical processes in fusion welding. Reviews of modern physics, 67(1):85112, 1995.

[2] Moore KL, Naidu DS, Ozcelik S. Modeling, sensing and control of gas metal arc welding. Kidlington (UK): Elsevier Science Ltd.; 2003.

[3] Kim, I.S.,Son, J.S.,Kim, I.G.,Kim, J.Y.,& Kim, O.S.A study on relationship between process variables and bead penetration for robotic CO2 arc welding. Journal of Materials Processing Technology, 136(1–3), 139–145,2003.

[4] Kokabi, A.H., & Ghaznavi, M.M. Welding technology [In Persian], First Volume. Sharif University of Technology Publication;2005.

[5] Doodman Tipi, A.R. The study on the drop detachment for automatic pipeline GMAW system: Free flight mode. International Journal of Advanced Manufacturing Technology, 50, 149–161,2010.

[6] Chu, Y.X., Hu, S.J., Hou, W.K.,Wang, P.C., & Marin, S.P .Signature analysis for quality monitoring in short-circuit GMAW. Welding Journal, 83(12), 336–343,2004.

[7] Kim, I.S., Jeong, Y.J., Son, I.J., Kim, I.J., Kim, J.Y., & Kim, I.K., et al. Sensitivity analysis for process parameters influencing weld quality in robotic GMA welding process. Journal of Materials Processing Technology, 140(1–3),676–681,2003.

[8] Mousavi Anzehaee, M., & Haeri, M.Estimation and control of droplet size and frequency in projected spray mode of a gas metal arc welding (GMAW) process. ISA Transactions, 53(3), 408–417,2011.

[9] Thomsen JS. Advanced control methods for optimization of arc welding. Ph.D. thesis. Denmark: Department of Control Engineering, Institute of Electronic Systems, Aalborg University; 2004.

[10] Abdelrahman, M.A. Feedback linearization control of current and arc length in GMAW systems', Proc. American Control Conference, Vol. 3, pp.1757–1761.1998.

[11] Moore, K.L., Abdelrahman, M.A. and Naidu, D.S. Gas metal arc welding control-II. Control strategy', Nonlinear Analysis, Vol. 35, No. 1, pp.85–93.1999.

[12] Kharaajoo, M.J., Gholampour, V., Ebrahimirad, H. andAshari, A.E. Robust nonlinear control of current and arc length in GMAW systems', Proc. Conference on Control Applications, Vol. 2, pp.1313–1316.2003.

[13] Thomsen, J.S. Control of pulsed gas metal arc welding, Int. J. Modelling, Identification and Control, Vol. 1, No. 2, pp.115–125.2006.

[14] Khatamianfar, A., Fateh, M.M. and Farahani, S.S. On sliding mode control of the manual gas metal arc welding process', IEEE Int. Conf. on Systems, Man and Cybernetics, pp.3570–3575.2008.

[15] Golob.M. Integrated Models of a Gas Metal ARC Welding Process and Inverter based Power Supply for Process Control Simulation Studies. ELEKTRONIKA IR ELEKTROTECHNIKA, VOL. 20, NO. 7,2014.

[16] Ozcelik, S., Moore, K.L. and Naidu, S.D, " Application of MIMO direct adaptive control to gas metal arc welding , " Proceeding of the American Control Conference, Philadelphia, USA,1998.

[17] Moore, K.L., Naidu, D.S., Abdelrahman, M.A. and Yesildirek, A, Advance welding control project, Annual Report FY96. Technical Report, ISU, Pocatello,ID, USA , 1996.

[18] Shepard,M. E. and Cook,G. E, A frequency-domain model of self-regulation in gas-metal arc welding, In Third International Conference on Trends in Welding Research, 899-903,1992.

[19] Wu, C., Chen, M. and Li, S.K, Analysis of droplet oscillation and detachment in active control of metal transfer, Computational Materials Science, 31,147–154, 2004.

[20] Arif, N, Lee, JH, Yoo, CD, Modeling of globular transfer considering momentum flux in GMAW, Journal of Physics D: Applied Physics, 41, 1–6, 2008.

[21] Rehan, M, Hong, KS, Ge, SS, Stabilization and tracking control for a class of nonlinear systems, Nonlinear Anal. Real World Appl. 12: 1786-1796,2011.

ISSN 1737-9334

Solving the Sensitivity Dependency Problem of Bio/Gas Sensor for Early Diagnosis and Life Critical Gas Detection

Hasan Göktaş

Electrical and Electronic Engineering, Harran University
Şanlıurfa 6300, Turkey
hgoktas@harran.edu.tr

***Abstract*— Reliability and the repeatability are the most important features for portable biosensor and gas sensors and any adverse effect (temperature, etc.) on these features must be studied in depth for the sake of high accuracy and long-term stable operation. Here, temperature effect on the sensitivity of the bio/gas sensor was studied in detail via FEM analysis. A 1 ng mass was added to the top surface of the clamped-clamped beam to observe the effect of the temperature on the sensitivity. Up to 10-fold (138/13.5) decrease in the sensitivity was observed when aluminum was used according to FEM results. The optimization in dimension decreased the 10-fold to 1.5-fold, whereas building clamped-clamped beam from silicon solved the problem by offering a negligible change (5%) in the sensitivity. These findings can be used to design more reliable and accurate bio/gas sensor for early cancer treatment and life critical gas detection while eliminating the need for compensation circuit.**

***Keywords*— biosensors, gas sensor, clamped-clamped beam, sensitivity, mass detection, temperature effect, frequency shift**

I. Introduction

Resonators have been widely used in building biosensors [1], gas sensor [2], temperature sensors [3] and thermal detectors [4]. Highly sensitive (ng-ag), low cost and portable bios/gas sensors are the most popular ones among the other applications in the last decades. Many different approaches have been proposed to build micro/nano resonators such as SAW [5], Cantilever [2], Fixed-Fixed beam [6], PZT [7], Whispering gallery mode [8], photonic crystal nano-ring resonator [9]. Moreover, micro/nano mechanical biosensors such as cantilever has many advantages in comparison to other types [10].

Many studies have been conducted on accuracy, large scale integration [11], sensitivity [12], selectivity and reliability in realizing bio/gas sensors. Reliability among other parameters keep a special place, because bio/gas sensor must give the same result no matter what the temperature in the environment is. High reliability is also very crucial for the early cancer diagnosis as it has a tremendous potential to cure the patients. Different calibration techniques such as compensation circuit with a programmable CMOS amplifier was proposed [13] to address the temperature effect.

Here, we study and propose solutions for the effect of temperature change on the sensitivity of the Clamped-Clamped beam type bio/gas sensor. The sensitivity was drastically decreased with even small increase in the temperate (14 Celsius) when aluminum was used to build the sensor. Up to 10-fold decrease in the sensitivity was observed according to FEM simulation results. The sensitivity was decreased from 138 kHz/1 ng to 13.5 kHz/1 ng when the temperature increased from 20 to 34 Celsius. Different techniques were used to address the problem. Optimizing length decreased 10-fold to 1.7-fold while optimizing thickness decreased it to 1.5-fold. However, this improvement is not enough for highly sensitive sensor and design still requires the compensation circuit [13] to calibrate the results. On the other hand, almost no decrease (5% change) in the sensitivity was observed when silicon was used to build the clamped-clamped beam type sensor. This is attributed to the fact that silicon has smaller thermal expansion constant in contrast to the aluminum layer. Using silicon can also eliminate the need for the compensation circuit and design complexity.

II. Results and Discussions

A 100 µm long, 1 µm thick and 4.5 µm wide aluminum clamped-clamped beam was modelled and built in COMSOL. The model can be easily fabricated with three deposition and etching processes. The first metal layer can be deposited via chemical vapor deposition (CVD) process and can be patterned via lithography and deep reactive ion etching process (DRIE). The isolation layer can be deposited on top of the bottom metal via atomic layer deposition (ALD) process. The top metal layer can be deposited on top of the isolation layer via CVD and patterned via DRIE process. The final step would be to remove the sacrificial layer so that the clamped-clamped beams can be released and free to resonate. Here sacrificial layer is not only used to separate the top and bottom metal during the fabrication process, but also to enable top metal etching without hurting the bottom metal.

The model was placed in a vacuum with a room temperature of 293 K. Solid mechanics was used to define the vacuum and ambient temperature. Two different structures were defined to study the sensitivity. The first one was defined as a simple clamped-clamped beam while the second structure was defined as a clamped-clamped beam with a load (1 ng mass) placed on top of it with a uniform distribution. A fine tetrahedral mesh was defined for both structures in the

ISSN 1737-9334

simulation environment. The eigenfrequency and modal response for both structures were calculated via COMSOL where the ambient temperature was swept from 20 Celsius to 34 Celsius. The purpose here is to find the resonance frequency for each case and consequently to derive the how sensitivity changes with respect to increase in the temperature. Here 1 ng mass represents biomolecules for biosensor application or gas molecules for gas sensor application.

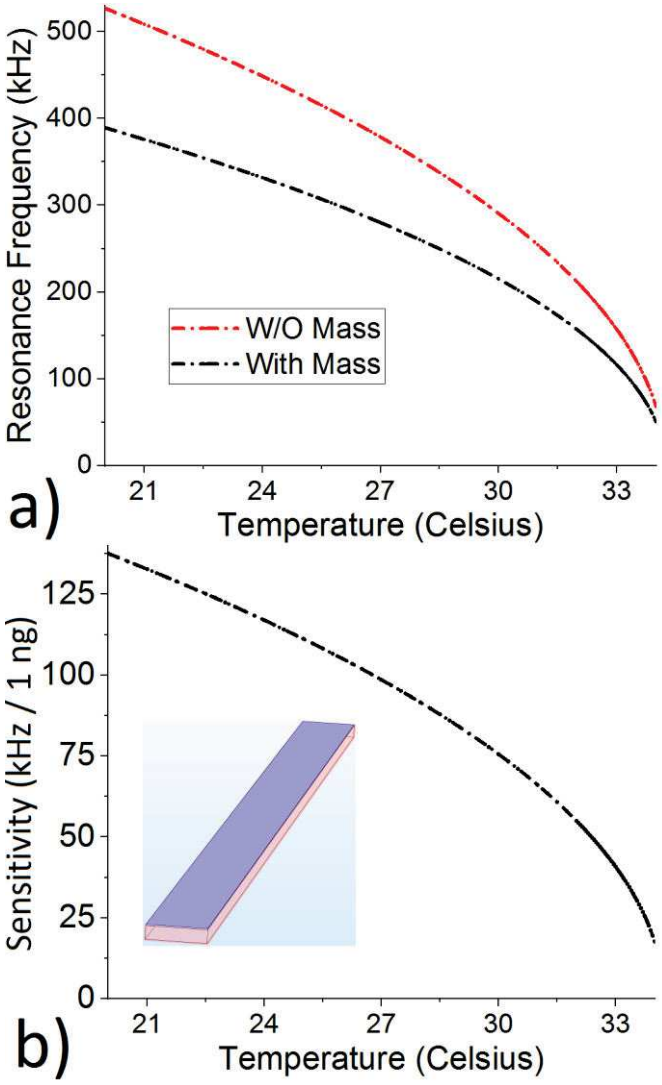


Fig. 1 a) Resonance frequency with respect to temperature for both aluminum clamped-clamped beam with mass (1 ng) and without mass, b) The effect of temperature on the sensitivity, inset is showing the aluminum clamped-clamped beam (red) with a mass (blue) attached on it.

The sensitivity can be found from the frequency shift with respect to the attached mass, where frequency shift is given in [14] as;

$$\frac{0.48\rho Lwh}{f_o}$$

Where w is the width, h is the thickness, ρ is the density and L is the length.

The resonance frequency decreases with the increase in the temperature due to decrease in the stiffness constant. The resonance frequency decreased from 527 kHz to 52 kHz for the aluminum clamped-clamped beam before a 1 ng mass and it decreased from 389 kHz to 38.5 kHz after adding a total of 1 ng uniformly distributed mass (Fig. 1a). This represent a total frequency change of 475 kHz for the design without a mass and 350.5 kHz for the design with a mass, where the total temperature change is 14 Celsius. Here the sensitivity is the difference of the resonance frequencies of the clamped-clamped beams without a mass and with a mass (Fig. 1a). It was observed that the sensitivity of the clamped-clamped type bio/gas sensor depends on the temperature and decreases with the increase in the temperature. The sensitivity was 138 kHz/1 ng (527-389=138) around 20 Celsius and it decreased to 13.5 kHz/1 ng (52-38.5) around 34 Celsius (Fig. 1b). This represents around 10-fold decrease in the sensitivity for the bio/gas sensor.

Two different approaches were studied in this work for the sake of not to analyse the problem but also to address the problem. First approach (dimensional effect) was to change the length (Fig. 2a) or the thickness (Fig. 2b) or both for the aluminum clamped-clamped beam. Three different lengths were studied via FEM. The linear relationship between sensitivity and the temperature was observed for all devices. Total loss in the sensitivity is around 119.8 kHz for 100 µm long beam, it is 99 kHz for 90 um long beam and 96.4 kHz for 80 µm long beam where temperature increased from 20 Celsius to 34 Celsius. A 100 µm long beam results in 10-fold decrease in the sensitivity while it was 2.6-fold for 90 µm long beam and was 1.7-fold for 80 µm long beam (Fig. 2a). In another words, the sensitivity dependency on the temperature can be decreased from 10-fold to 1.7-fold by simply changing the length. Though the total sensitivity change (96-120 kHz) with respect to temperature change is very close to each other for all devices, the relatively smaller sensitivity (137.6 kHz) of 100 µm long beam in comparison to 80 µm long beam (248 kHz) at 20 Celsius results in 10-fold decrease.

In contrast to the length study, the thickness study shows different slopes and curves for devices at different thickness. The total sensitivity change with respect to the total temperature change of 14 Celsius was calculated as 119.8 kHz for 1 µm, while it was 63 kHz for 1.2 µm and 44.6 kHz for 1.4 µm according to COMSOL (Figure 2b). A 10-fold decrease in the sensitivity was observed for 1 µm thickness, while it is 1.9-fold for 1.2 µm thick beam and 1.5-fold for 1.4 µm thick beam (Fig. 2b). Although each device has a sensitivity around between 137-153 kHz at 20 Celsius, the thinner device (1um) has a curve with a larger slope that results in 119.8 kHz change in the sensitivity.

These findings can be attributed the fact that, the increase in the stiffness constant results in smaller percentage decrease in the sensitivity with respect to temperature. Here the decrease in the length or the increase in the thickness causes the increase in the stiffness constant. These results can also be explained by the temperature sensitivity. Thinner or longer devices are more sensitive to the temperature change [3] that

ISSN 1737-9334

adversely affect the performance in biosensing applications. Therefore, the increase in the thickness or decrease in the length is desirable if one wants to decrease the temperature sensitivity and consequently decrease the temperature dependency of the biosensor. Another way is to build a package with thermal isolation; however, this solution method is not practical and realistic for the biosensor applications because biosensors need microchannels (mostly build on PDMS) on top of the devices and that would prevent the thermal isolation. If somehow one design a microchannel built on different material with a good thermal isolation, that still would bring the fabrication complexity, and cost

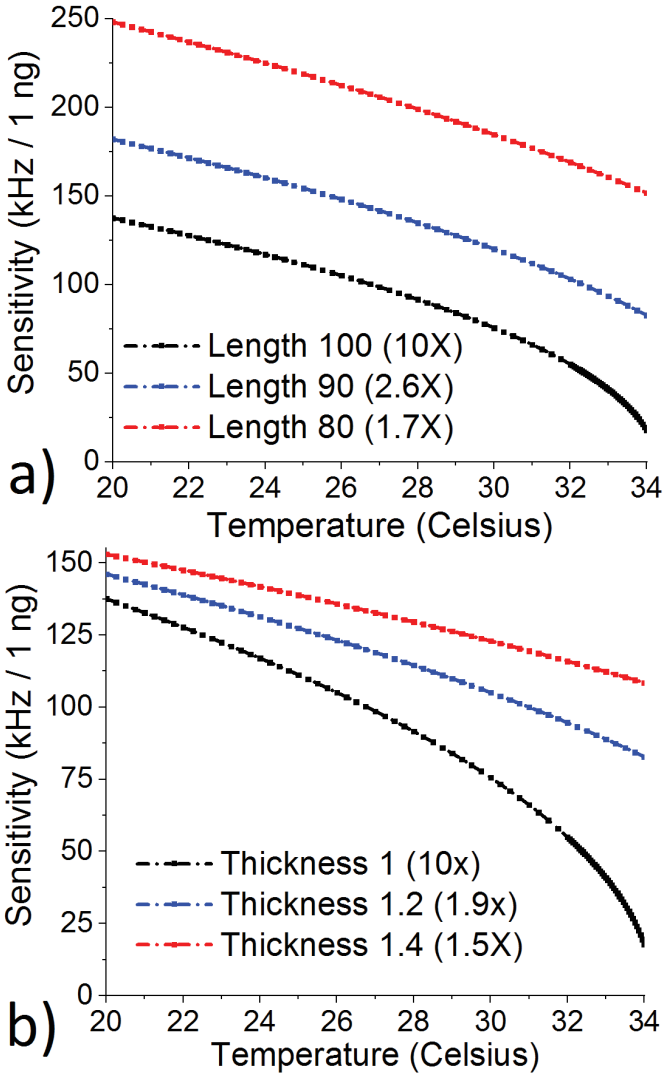


Fig. 2 The effect of temperature on the sensitivity for an aluminum clamped-clamped beam with a) a length of 100 µm, 90 µm and 80 µm where thickness is 1 µm, b) a thickness of 1 µm, 1.2 µm and 1.4 µm where length is 100 µm

Tough first approach partially solved the problem and causes a smaller percentage decrease in the sensitivity with respect to temperature, a 1.5-fold can still not acceptable for reliable and accurate detection especially for the application such as early diagnosis, where the sensitivity requirement can reach to atto-gram. That's why, another simple and cost-effective solution is required to address the problem.

The second approach was to change the material in building the clamped-clamped beam. This can be done easily by switching to SOI (Silicon on insulator) process and taking the advantage of smaller thermal expansion constant of silicon. SOI is a widely used process and some examples can be given as MEMS-resonators [13], and plasmonic switches [15]. The structure consists of Silicon layer growth on oxide with different thickness depending on the application. The device can be built by first patterning the silicon layer and then etching the oxide layer to release the devices.

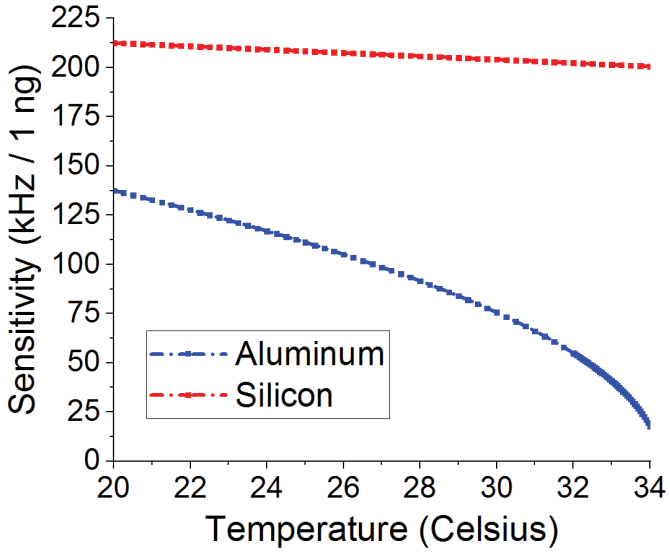


Fig. 3 The effect of temperature on the sensitivity for an aluminium clamped-clamped beam and silicon clamped-clamped beam

The effect of temperature on the sensitivity for the clamped-clamped beam built on silicon and aluminum were studied via FEM (Fig. 3) where the beam is 100 µm long, and 1 µm thick. The sensitivity decreased from 212 kHz to 200 kHz for silicon and 137.6 kHz to 17.8 kHz for aluminium type beam, where the temperature was increased from 20 to 34 Celsius. In another words, the total decrease in the sensitivity is 12 kHz for silicon beam while it is 119.8 kHz for aluminum. This represent around 5.6% change in the sensitivity for silicon beam, where ΔT is 14 (34-20), thanks to smaller thermal expansion constant of silicon in comparison to aluminum ($\alpha_{aluminum}$ = 23x10^{-6} m/mK, $\alpha_{silicon}$ = 3-5x10^{-6} m/mK). This 5.6% change is a drastic increase and serious improvement over aluminum beam where the sensitivity was decreased by 10-fold.

III. Conclusions

Clamped-clamped beams provide high sensitivity in portable biosensor and gas sensor applications, that's why it keeps a special and critical place among other micro devices. The accuracy and reliability for portable biosensor and gas sensor are the most important features as they provide important information for early diagnosis for cancer treatment and life critical gas sensing. 10-fold decrease in the sensitivity was calculated via FEM for the clamped-clamped beam built on aluminum, where the total temperature difference (ΔT) was 14 Celsius. Changing the length or the thickness decrease

ISSN 1737-9334

the 10-fold down to 1.5-fold, however this is still a large value for the highly sensitive bio/gas sensor. On the other hand, the solution for this problem was achieved by changing the material from aluminum to silicon while eliminating the need for any change in the dimensions. The sensitivity was only changed from 212 kHz to 200 kHz (5%) for the clamped-clamped beam built on silicon. This small change in the sensitivity with respect to temperature is negligible in comparison to 1.5-fold achieved by changing the thickness of aluminum beam.

REFERENCES

[1] M. G. von Muhlen, N. R. Brault, S. M. Knudsen, S. Jiang, and S. R. Manalis, "Label-Free Biomarker Sensing in Undiluted Serum with Suspended Microchannel Resonators," *ACS Anal Chem*., vol. 82, pp. 1905-1910, Feb. 2010

[2] N. Kilinc, O. Cakmak, A. Kosemen, E. Ermek, S. Ozturk, Y. Yerli, Z. Z. Ozturk, H. Urey, "Fabrication of 1D ZnO nanostructures on MEMS cantilever for VOC sensor application," *Sensors and Actuators B: Chemical*, vol. 202, pp. 357-364, October 2014

[3] H. Göktaş, K. L. Turner, and M. E. Zaghloul, "Enhancement in CMOS-MEMS Resonator for High Sensitive Temperature Sensing," *IEEE Sensors Journal*, vol. 17, pp. 598-603, Feb. 2017.

[4] H. Göktaş, "Towards an Ultra-Sensitive Temperature Sensor for Uncooled Infrared Sensing in CMOS–MEMS Technology," *Micromachines*, vol. 10, pp. 1-7, Feb 2019

[5] A. Mujahid and F. L. Dickert, "Surface AcousticWave (SAW) for Chemical Sensing Applications of Recognition Layers," *Sensors*, vol. 17, pp. 1-26, November 2017

[6] H. Göktaş and M. E. Zaghloul, "Tuning In-Plane Fixed–Fixed Beam Resonators With Embedded Heater in CMOS Technology," *IEEE Electron Device Letters*, vol. 36, pp. 189-191, Feb. 2015

[7] D. G. Hwang, Y. M. Chae, K. S. Hwang, J. Y. Kang, and S. H. Lee, "Fabrication and characterization of PZT (lead zirconate titanate) bridge-shaped resonator for mass sensing application," *Journal of Electroceramics*, vol. 29, pp. 225-234, November 2012

[8] M. R. Foreman, J. D. Swaim, and F. Vollmer, "Whispering gallery mode sensors," *Advances in Optics and Photonics*, vol. 7, pp. 168-240, May 2015

[9] S. Olyaee, and A. Mohebzadeh-Bahabady, Designing a novel photonic crystal nano-ring resonator for biosensor application," *Optical and Quantum Electronics*, vol. 47, pp. 1881-1888, July 2015

[10] J.L. Arlett, E.B. Myers and M.L. Roukes, "Comparative advantages of mechanical biosensors," *Nature Nanotechnology*, vol. 6, pp. 203-215, March 2011

[11] I. Bargatin, E. B. Myers, J. S. Aldridge, C. Marcoux, P. Brianceau, L. Duraffourg, E. Colinet, S. Hentz, P. Andreucci, and M. L. Roukes, "Large-Scale Integration of Nanoelectromechanical Systems for Gas Sensing Applications," *ACS NanoLetters*, vol 12, pp. 1269-1274, January 2012

[12] Y. T. Yang, C. Callegari, X. L. Feng, K. L. Ekinci, and M. L. Roukes, "Zeptogram-Scale Nanomechanical Mass Sensing," *ACS NanoLetters*, vol. 6, pp. 583-586, March 2015

[13] M. S. Islam, R. Wei, J. Lee, Y. Xie, S. Mandal, and P. X.-L. Feng, "A Temperature-Compensated Single-Crystal Silicon-on-Insulator (SOI) MEMS Oscillator with a CMOS Amplifier Chip" Micromachines, vol. 9, pp. 1-13, September 2018

[14] G. M. Rebeiz, *RF MEMS Theory, Design and Technology*, John Wiley & Sons, 2003.

[15] H. Göktaş, Fikri Serdar Gökhan, and Volker J. Sorger, "Electrical-Driven Plasmon Source of Silicon Based Quantum Tunneling" *ACS Phootnics*, Vol. 5 pp 4928-4936, November 2018

ISSN 1737-9334

Human Activity Recognition based Machine Learning for Multimodal Robot Interaction

Souhila Kahlouche*, Mahmoud Belhocine**

*Ecole nationale Supérieure de l'Informatique ESI, OuedSmar, Algiers, Algeria
**Centre de Développement des Technologies avancées CDTA, Baba Hassen, Algiers, Algeria

S_kahlouche@esi.dz
mbelhocine@cdta.dz

***Abstract*--The ability to recognize human activities is necessary to allow intuitive Human Robot Interaction (HRI) for a service robot in real time.**
In this work, a HRI system for a service robot capable of understanding common interactive human activities is developed. The use of supervised machine learning algorithms for the human activity recognition task is based on the fusion of several classifiers like Support Vector Machine (SVM), Decision Tree (DT) and Artificial Neural Network (ANN) algorithms to classify activities into different classes. It uses as input a view invariant 3D data of the skeleton joints, which are rich body movement information recorded from a single Microsoft Kinect camera to create specific dataset of six interactive activities.
The algorithm was able to successfully classify and recognize the activities being performed in front of the camera. The system framework is realized on the Robot Operating System (ROS), and the real-life activity interaction between our service robot and the user was conducted to demonstrate the effectiveness of thedeveloped HRI system.

***Keywords*— Human Robot Interaction (HRI), Human Activities Recognition (HAR), Machine Learning, Robot Operating System (ROS), Service Robot, RGB-D camera.**

I. INTRODUCTION

Interactive scenarios for social robots require not only the ability to find and track humans, but also to understand their gestures and actions.
Human Action Recognition (HAR) is a particularly important aspect of social interaction, which uses a combination of sensors and modalities, including vision, audio, and touch to recognize human actions. This aspect has to be considered in every Human Robot Interaction (HRI) system.
Machine learning algorithms have been widely applied for this purpose. Using Neural Network, Fiaz and Ijaz [1] proposed a method to design an intelligent human activity monitoring system to detect and track suspicious activity in a surveillance environment. A three-layer perceptron was used for the classification of the human activities from the information of the distance and motion vector. Foroughi *et al.* [2] applied four-layer MLPs to learn Eigen-motions, which are the movement patterns extracted by an Eigenspace technique, for motion classification and falling detection.
Using Support Vector algorithm, Schuldt *et al.* [3] aimed to recognize human activities by extracting local space-time features in a video. In addition, Laptev *et al.* [4] used a nonlinear SVM with a multi-dimensional Gaussian kernel for recognition of various natural human activities by building spatial-temporal bag-of-features (space-time grids).
Decision tree algorithm also constitutes one of the most used machine learning methods in HAR problem. Stauffer and Grimson [5] built a hierarchical classification binary tree by the accumulated joint co-occurrence statistics of the representations in a video sequence. Moreover, Ribeiro and Santos-Victor [6] investigated the use of hierarchical binary tree classifiers on human activity recognition. For each node of the binary tree classifier, a Bayesian classifier was used and the likelihood functions were modeled and systematically learned as Gaussian mixtures.
Recently, the availability of low-cost and high-accuracy depth sensors, using the out-of-the-box skeletal tracker framework [7]and[8]leads toa large community of researcher to investigate in motion and pose tracking fields. Anjum *et al.* [9] used a Kinect sensor as input for an activity recognition classifier, with a basic SVM classifier that achieved near perfect accuracy on a closed set of eight activities. McKeague*et al.* [10] used sensor fusion (depth and RGB) to track hands in crowded environments in real time. A Monte-Carlo update process reduced false positives of varying skin color, clothing, and illumination conditions.

ISSN 1737-9334

To encourage HRI-focused activity recognition, Chrungoo *et al.* [11] released an annotated RGBD human-robot interaction dataset consisting of 18 activities including ten stylized gestures and eight conventional activities of daily living. The dataset includes both communicative and non-communicative actions; they used basic SVM classifier that achieved near perfect accuracy on a closed set of eight activities.

In this work, we take into consideration that a set of classifiers is designed to obtain better performance than any of individual one. For this purpose, we combine three single classifiers: Support Vector Machine, Decision Tree and Neural Network to classify six interactive activities represented by 3D coordinate of skeleton data.

The major contributions of this work are:

- The proposed viewpoint invariant transformation in the preprocessing step, which can eliminate observation variations caused by viewpoint differences.
- Provide ROS package and a specific dataset containing interactive activities for the HRI system. This dataset is a challenging one, once there is significant intra-class variation in the realizations of the same activity. For example, sometimes 'Call' activity is done with the left hand while sometimes with the right hand.

II. HUMAN-ROBOT INTERACTION SYSTEM OVERVIEW

Figure 1 shows the framework of human-robot interaction system. It mainly includes two modules:

The off line process where the dataset is created using RGBD camera to obtain the human skeleton data stream, from which features are extracted and pretreated. Three supervised machine learning algorithms (SVM, DT, ANN) are then trained.

The online process where the activity being performed is inferred using voting process of the above classifier and the robot runs the action according to rules that are described in Table1.

III. THE OFF-LINE PROCESS

A. Dataset Creation

Our dataset has been created using *OpenNI* tracker framework, which allows the skeleton tracking at30 fps, providing the three-dimensional Euclidean coordinates and three Euler angles of rotation in the 3D space for each joint with respect to the sensor. The dataset contains six interactive activities performed by different individuals, among the activities. Four are static, where the user does not move from camera view field like (hello, Stop, call), and two dynamic activities where the user can leave the camera view field like (going, coming), figure2 shows some examples of our dataset.

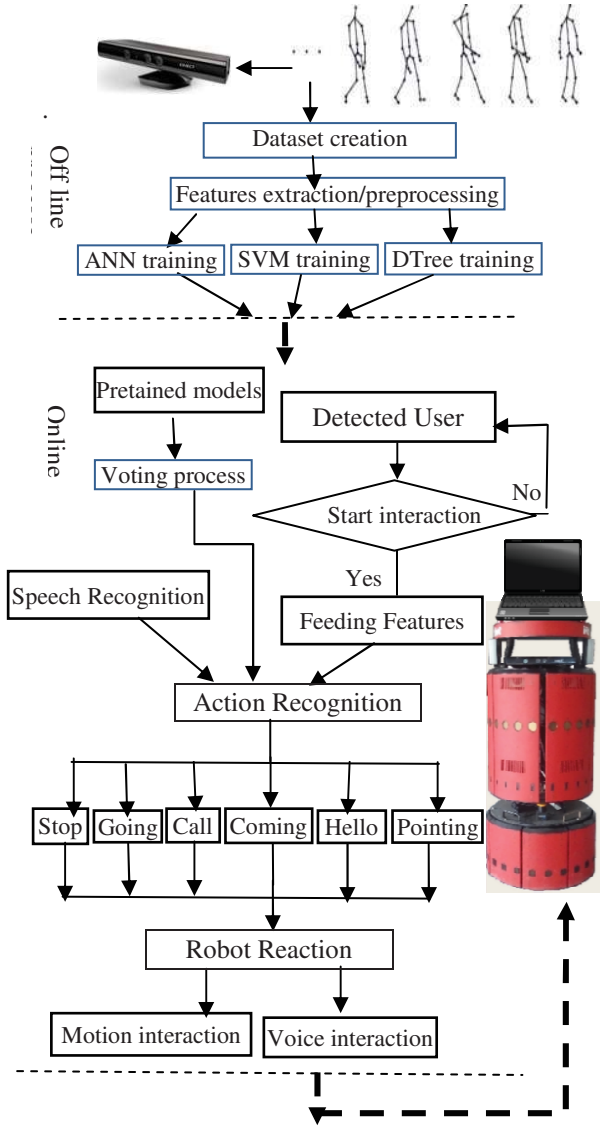


Fig.1: The framework of human-robot interaction system

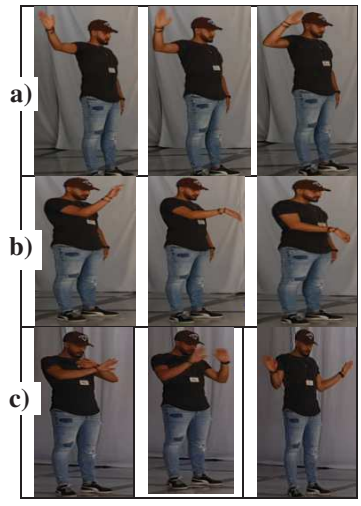


Fig.2: Examples of our dataset a) Hello, b) Call, c) Stop

ISSN 1737-9334

B. Features Extraction

In order to detect and recognize activities related to interactive service robot, we have selected nine joints in the upper body human skeleton (see figure3), since joints on low limbs introduce more noise than useful information to distinguish activities.

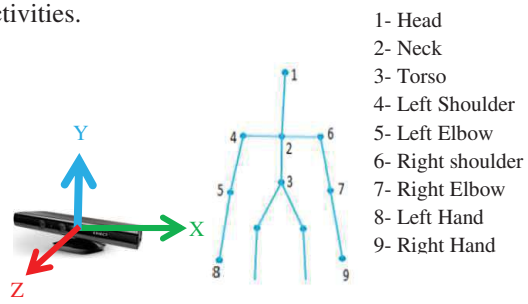


Fig.3: The selected human skeleton joints

1) Construction of Feature Vectors

Using the above information, we have compute dataset of features as follow:

- 3D coordinates of the 9 selected joints: (x,y,z).
- 3D rotation angles in the space described by a sequence of three rotations according to the three axes X, Y and Z. Let Ψ be the rotation angle about the x-axis computed as:

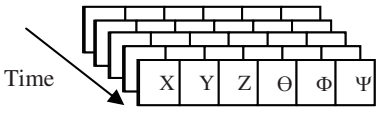
$$R_x(\Psi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\Psi & -sin\Psi \\ 0 & sin\Psi & cos\Psi \end{bmatrix} \quad (1)$$

Similarly, let Θ and Φ be the rotation angles about the y-axis and z axis respectively:

$$R_y(\Theta) = \begin{bmatrix} cos\Theta & 0 & sin\Theta \\ 0 & 1 & 0 \\ -sin\Theta & 0 & cos\Theta \end{bmatrix} \quad (2)$$

$$R_z(\Phi) = \begin{bmatrix} cos\Phi & -sin\Phi & 0 \\ sin\Phi & sin\Phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

Hence, the feature vector has the following form:



2) Feature preprocessing

A preprocessing step is applied to the 3D skeleton data, it consists of the following:

Translation: to guarantee the same origin of the coordinates system for all acquired frames; the reference was set to the torso of the human skeleton;

Smoothing: in order to attenuate noise introduced by the sensor, we have applied a moving average filter with five (5) neighbor's data points, to filter the noise and smooth the data.

Normalization: a normalization step is applied to accommodate and reduce the influence of different user's height and limb lengths; therefore, all skeleton 3D coordinates are normalized as follow;

$$D_{i\text{-normal}} = \frac{D_i - \min(D_i)}{\max(D_i) - \min(D_i)}$$

Where $\max(D_i)$ and $\min(D_i)$ are the maximum and minimum value of a feature set.

Symetrization: It is done to disambiguate between mirrored versions of the same activity (e.g. gestures performed by right and left-handed people); it is required for activities like Hello, Call and pointing. It is just necessary to consider a new sample based on a mirrored version of the original 3D skeleton data.

C. Training

An ensemble learning model is applied in order to obtain better prediction result than any individual classifiers [12]. This is done by using multiple learning algorithms; in this work, we have used a fusion of three single classifiers, which are non-linear support vector machine using the *X2* kernel function, Gaussian and polynomial kernel and Neural network using '*relu*' as activation function and Decision Tree(see figure 4).

In order to validate the activities recognition models, 70% of our dataset has been used for training while 30% for testing.

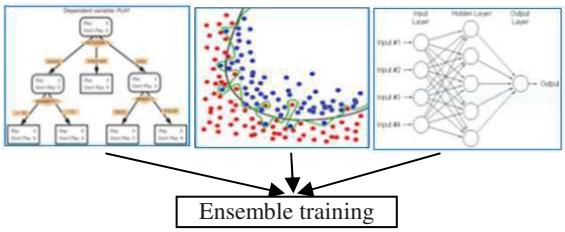


Fig.4: Ensemble learning

1) *Support Vector Machine:* is discriminative classifier, which outputs the optimum hyperplane that separates the training data. Margin is defined as the distance of data from the separating hyperplane. The classification results of SVM that are applied to our dataset are presented in a confusion matrix, in which the performance measures of accuracy is 84% (fig.5).

ISSN 1737-9334

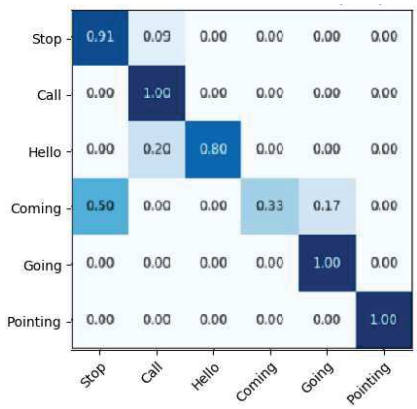


Fig.5: Confusion Matrix SVM

2) Decision Trees: The goal of Decision Trees is to generate a model that predicts the classification labels based on decision rules according to the training data. At each level of the tree, the model tries to maximize information gain by considering all possible splits based on the used features.

Binary tree classification is simple and fast, but the separation rules in each node are difficult to be general for other cases, making it difficult for complex scenarios.

The classification results are shown in figure 6 where the accuracy result is 94%.

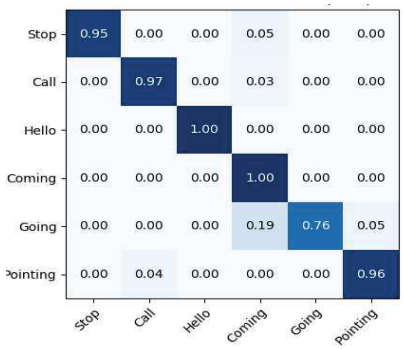


Fig.6: Confusion Matrix Decision Tree

3) Artificial Neural Networks use the concept of biological neuron networks to supply nonlinear capability via more discriminative feature transformation through hidden layers, resulting in more effective classification through multilayer perceptron (MLPs), which apply the back propagation (BP) supervised learning algorithms to compute suitable connection weights and biases of the network.ANN is applied on our dataset with '*relu'* as activation function and achieved 72.5% of accuracy (see figure 7).

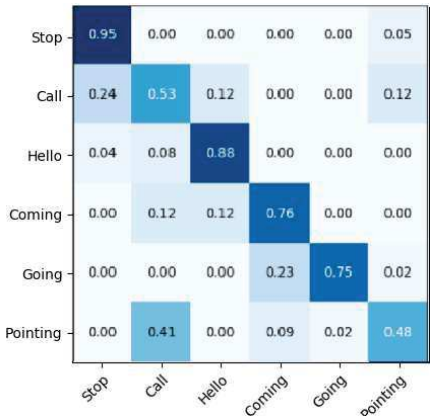


Fig.7: Confusion Matrix ANN

D. Voting process:

In general, ensemble learning involves training more than one model on the same dataset, then using each of the trained models to make a prediction before combining the predictions in some way to make a final outcome or prediction.

To combine multiple classifiers we have used majority-voting process as follows:

Let $M_i=\{m_1, m_2, \ldots, m_n\}$, be the ensemble of model trained and $C_j=\{c_1, c_2, \ldots, c_l\}$, be the classes.

The vote for each class j and each model i is computed by:

$$v_{i,j} = \begin{cases} 1 & if\ m_i\ votes\ for\ c_j \\ 0 & else \end{cases}$$

The total vote for each class j is then calculated by $V_j = \sum_{i=0}^{n} v_{i,j}\ ,\ j = 1, \ldots, l$

The final decision V_{final} is then the majority vote. $V_{final}=argmax(Vj)$.

IV. THE ONLINE PROCESS

A. Robot reaction

A multimodal interaction is designed for human robot interaction scenarios to provide natural communication. Once an activity is recognized by the classification step, the proper reaction is taken from the look-up table to be executed by the mobile robot.

ISSN 1737-9334

TABLE 1: Robot reactions

Activity	Robot reaction
Hello	*Voice reaction* : Hello welcome to CDTA
Call	*Voice reaction* : Please wait, I am coming *Motion reaction:* Approaching user and start interaction.
Stop	*Voice reaction:* ok, I will stop here. *Motion:* Stop moving.
Pointing	*Voice reaction:* Ok, I will go there. *Motion reaction*: Go to pointed position.
Coming	*Voice reaction*: How can I help you? *Motion reaction*: step back and prepare to begin interaction.
Going	*Voice reaction*: Good-bye, thank you for your visit. *Motion*: turn back and stop interaction

B. *Speech recognition module*

Speech recognition module is performed using one of the commercial softwares, which are embedded solutions. This choice is due to the autonomy issue of our service robot in order to make it light enough in terms of computational and memory resources, since other processes are running on the same platform and at the same time.

In our solution, we have used pocketsphinx [13] package to recognize a single word in order to confirm the activity being recognized and to make conversation between the robot and the human more natural. In our work, the vocabulary includes the following words:"no","yes", "please", thanks","goodbye".

C. *Speech synthesis module*

For voice reaction, we have used the sound play package that is given a text input, it will be synthesized into sound output.

V. RESULTS AND DISCUSSION

The developed HRI system has been implemented under Robot Operating System (ROS) and tested on the B21r mobile robot in real scenarios.

Our dataset contains several intra-class variations among different realizations of the same activity. For example, hello wave is done with the left or right hand. Another challenging feature is that the activity sequences are registered from different views.

The validation technique adopted for activity recognition was the leave-one-out cross-validation (LOOCV). The idea is to verify the capacity of generalization of the proposed classifier by using the strategy of learning from different persons and testing with an unseen person.

Figure 8 shows the confusion matrix of activity recognition using voting process, which achieves accuracy of 97.1% (see Table2).

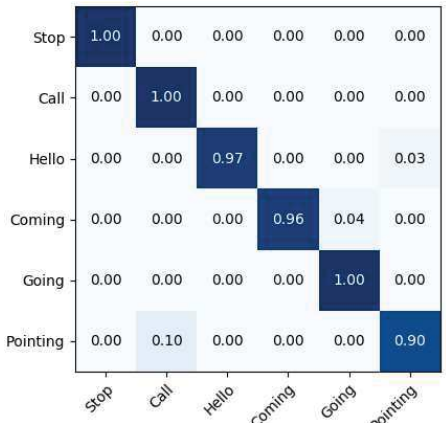


Fig.8: Confusion Matrix of voting process

From Table2, we can see that activity recognition accuracy of the fusion approach is much better than the base classifiers one. Therefore it has been used for the real time application.

Table2: Comparison of fusion accuracy with base classifiers

Method	Accuracy
SVM	84%
ANN	72.5%
DTree	94%
Fusion	**97.1%**

The experimental tests using the voting process approach for a real time application acquire 4 seconds of RGBD sensor data for features extraction and classification. After that, a final decision is made for activity recognition and a robot reaction is performed.

The proposed framework was capable of recognizing different interactive activity that happens sequentially in case of a person transit from one activity to another one.

Figure 9 shows a scenario where a person attempts to interact with a service robot.

a) First, the person salute the robot, 'Hello' action is recognized and the robot reacts with voice mode and says: '*Hello, welcome to CDTA'*.

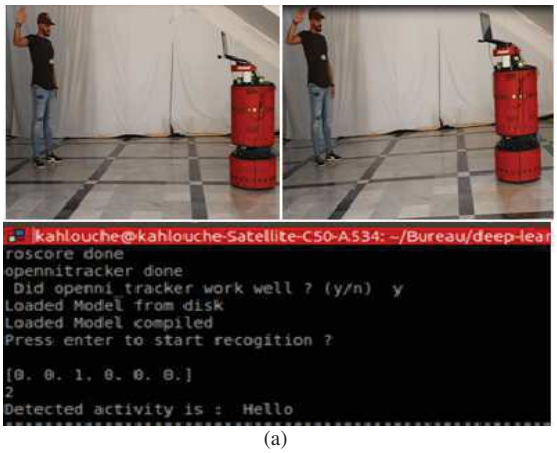
b) The person is calling the robot, and the activity is recognized as 'Call' and then the robot interacts with voice mode, and says: "*Please wait, I am coming*", at the same time approaching the user.

ISSN 1737-9334

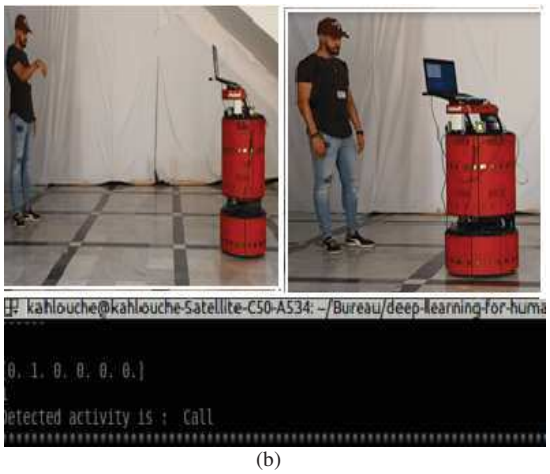
c) The user is pointing position; the activity is recognized as "pointing", and the voice reaction of the robot is: "*Ok, I will go there*", and the robot moves to the pointed position.

d) The user decides to stop the robot; the activity is recognized as 'Stop', and the robot stops moving.

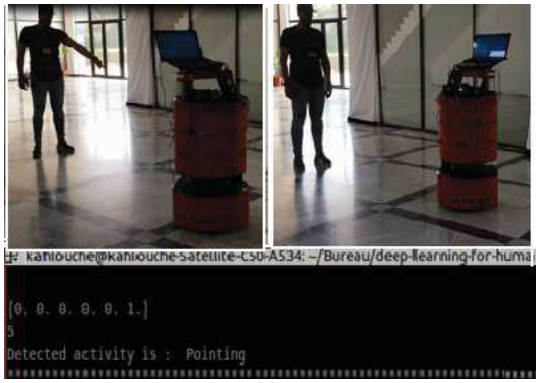
e) The user decides to go away, the activity "Going" is recognized and the robot reacts with voice: "Good-bye, thank you for your visit", and turns back and stops interaction.



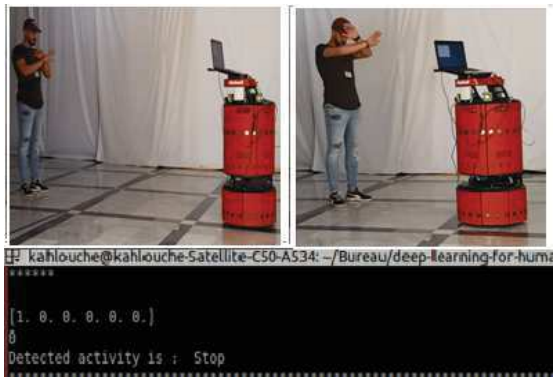
(a)



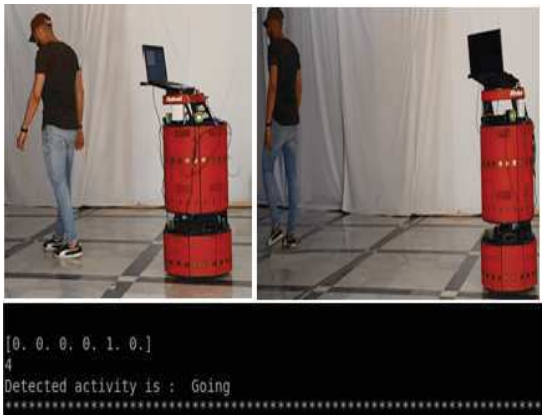
(b)



(c)



(d)



(e)

Fig.9: Human Robot Interaction scenario

VI. CONCLUSION

This work demonstrates progress towards a HRI system for a service robot, capable of understanding interactive human activities in real time.

In the preprocessing step, we have proposed view invariant transformation applied to the data, captured by the Microsoft Kinect camera to grantee view invariant features.

Three machine-learning algorithms which are SVM, Decision Tree, and ANN, have been used in this work and their classification results are compared. Based on those classifiers, a fusion approach has been proposed using voting process, and the experimentation shows that they are enough for obtaining good results.

Our HRI system was able to correctly recognize the performed activity and it successfully reacts accordingly to the situations.

However, the same activity may have different behavior interpretations depending on the context in which it is performed. More specifically, the time (when), the place (where), the intention (why), the manner (how). In the future, we will focus on the study of that additional contextual information,

ISSN 1737-9334

which will improve the interactivity of our service robot.

REFERENCES

[1]Fiaz, M.K.; Ijaz, B. Vision based Human Activity Tracking using Artificial Neural Networks. In Proceedings of IEEE International Conference on Intelligent and Advanced Systems (ICIAS), Kuala Lumpur, Malaysia, 15–17 June 2010; pp. 1–5.

[2] Foroughi, H.; Naseri, A; Saberi, A.; Yazdi, H.S. An Eigenspace-based Approach for Human Fall Detection Using Integrated Time Motion Image and Neural Network. In Procedings of IEEE 9th International Conference on Signal Processing (ICSP), Beijing, China, 26–29 October 2008; pp. 1499–1503.

[3] Schuldt, C.; Laptev, I.; Caputo, B. Recognizing Human Actions: A Local SVM Approach. In Proceedings of the 17th IEEE International Conference on Pattern Recognition (ICPR), Cambridge, UK, 23–26 August 2004; Volume 3, pp. 32–36.

[4] Laptev, I.; Marszalek, M.; Schmid, C.; Rozenfeld, B. Learning Realistic Human Actions from Movies. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.

[5] Stauffer, C.; Grimson, W.E.L. Learning patterns of activity using real-time tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 747–757.

[6] Ribeiro, P.C.; Santos-Victor, J. Human Activity Recognition from Video: Modeling, Feature Selection and Classification Architecture. In Proceedings of the International Workshop on Human Activity Recognition and Modelling (HAREM), Oxford, UK, 9 September 2005; Volume 1, pp. 61–70.

[7] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore,A. Kipman, A. Blake, Real-Time Human Pose Recognition in Parts fromSingle Depth Images, in: IEEE Conference on Computer Vision andPattern Recognition, 2011.

[8] NITE, http://openni.ru/files/nite/ (2012).

[9] M. L. Anjum, O. Ahmad, S. Rosa, J. Yin, and B. recognition using kinect camera. In SocialBona. Skeleton tracking based complex human activity Robotics, pages 23–33. Springer, 2014.

[10] S. McKeague, J. Liu, and G.-Z. Yang. An asynchronous rgb-d sensor fusion framework using monte-carlo methods for hand tracking on a mobile robotin crowded environments. In Social Robotics,pages 491–500. Springer,2013

[11] A. Chrungoo, S. S. Manimaran, and B. Ravindran. Activity recognition for natural human robot interaction. In *Social Robotics*, pages 84–94. Springer,2014.

[12] Zhi-Hua Zhou, “Ensemble Methods: Foundations and Algorithms”, CRC Press, 2012

[13] Walker W, Lamere P, Kwok P, et al. (2004) Sphinx-4: A flexible open source framework for speech recognition. *SMLI TR- 2004-139*, Mountain View, CA, USA: Sun Microsystems, Inc.,Available from: http://dl.acm.org/citation.cfm?id=1698193

ISSN 1737-9334

6th International Conference on Automation, Control, Engineering & Computer Science (ACECS-2019)
Proceedings of Engineering and Technology PET

Effect of Node Placement Strategies on Network Lifetime in Heterogeneous WSN

Yousef Jaradat*, Mohammad Masoud*, Ismael Jannoud* and Amer Zerek†
*Department of Electrical Engineering/Communications and Computer
Al-Zaytoonah University of Jordan - Amman, Jordan
Email: y.jaradat@zuj.edu.jo
†Department of Electrical Engineering, Zawia University, Zawia, Libya

***Abstract*—In this paper, the effect of different node placement strategies on maximizing the lifetime of heterogeneous wireless sensor network is investigated. Normal, exponential and uniform node placement distributions are utilized. Stable election protocol (SEP) is used to evaluate the performance of different node location distributions in terms of stability and instability periods, network lifetime and throughput. It was noticed that different node distributions have almost the same stable and unstable regions. It was also shown that normal distribution has a slightly larger throughput compared to the uniform and exponential distributions.**

***Index Terms*—Wireless Sensors Networks (WSN); SEP; MLE; Heterogeneous WSN**

I. Introduction

Wireless Sensor Network (WSN) are network of small battary-powered sensor nodes [1]. Typical sensor nodes are deployed randomly and uniformly in a target field. The purpose of sensor nodes is to sense the region in which they were deployed in. Moreover, the sensed data is then processed locally and transmitted to centralized processing node usually called the sink or the base station (BS). Data communication is one of the major functions done by a sensor node. Usually, data communication consume a huge amount of energy compared to other tasks done by a sensor node [2]. However, the design of WSN communication protocols and applications has to be energy-aware and efficient to extend the life span of the network since the replacement of energy sources in these nodes are hard once they deployed.

WSNs are used everywhere these days. They can be used in agriculture, mines, homes, cities, commerce, military and a variety of monitoring applications [3]. WSN nodes usually are deployed in large-scale depending on the application of interest [4].

Clustering algorithms are used intensively in routing data to the sink node. The importance of clustering algorithms in WSNs stems from the fact that they are energy efficient and extending the lifetime of the network by load balancing energy expenditure across all sensor nodes. The basic operation of clustering protocols is to divide sensor nodes into clusters. Every cluster consists of a number of member nodes (MN) and a cluster head (CH). MNs send their data to the CH and CH aggregates, compresses and transmit data to the BS directly or through multi-hop approach [5]. The formation of clusters and the election process of CH are done continuously in protocol rounds. All nodes are becoming CHs in different rounds to balance energies between them and to prevent quick nodes death. An epoch is defined as the sequence of rounds in which all network nodes become CHs. In this way, energy expenditure is distributed equally among different epochs and as a result network lifetime is prolonged. Low energy adaptive clustering hierarchy (LEACH) [6] and stable election protocol (SEP) [7] are typical examples of clustering algorithms.

Two types of WSNs are usually studied: Homogeneous and heterogeneous. In homogeneous WSN (HO-WSN) sensor nodes are identical in their structure, sensors, computation and energy [8]. In heterogeneous WSN (HE-WSN) nodes have different capabilities, structure, computation and energy [7]. In this paper heterogeneity is limited to the energy source. A percentage of nodes (m) is supplied with more energy than others. Nodes with more energy than others are referred to as advanced nodes, other nodes are referred to as normal nodes.

In many research studies, nodes' locations are distributed uniformly in HO-WSN or HE-WSN [9], [10]. No location is preferred than another. All locations in the network field are likely to have a sensor node. In some network settings nodes are distributed according to other distributions. This mainly depends on the application of WSN, terrain of the field and the required quality of service. In this paper, normal and exponential node distributions a long with the uniform distribution are used to study the impact of different node distributions on energy expenditure, network lifetime of HE-WSNs.

The rest of the paper is organized as the follows. Section II overviews the related work. Section III describes models and assumptions used in this study. Simulation settings and results are provided in section IV. Finally, we conclude our research in section V.

II. Related Work

A number of research studies have been conducted regarding node deployment strategies in WSN environment. The authors in [11] studied maximization of WSN lifetime per unit cost, this metric is defined as the WSN lifetime divided by the number of network nodes. In [12] the authors formulate a non-linear optimization problem in order to determine the optimal nodes' distributions for efficient energy utilization. The authors in [13] showed that in hard-terrain areas the probabilistic

ISSN 1737-9334

distribution of nodes is better than the uniform distribution in terms of WSN intruder detection, network lifetime and stability of routing algorithms. In [14] the authors proved analytically that normal distribution of sensor nodes improves intruder detection probability for a target field. The authors in [15] carried out a comparative simulation-based study for node deployment strategies using uniform, normal and mixed distributions. Their findings indicated that normal distribution of nodes is better than the others in terms of average number of hops, intrusion detection and the average number of control bits. On the other hand mixed node distribution showed better results in reducing end-to-end packet delay compared to other distributions. In [16] the author showed that energy efficient routing protocols for maximizing the lifetime of WSN are improved utilizing node placement strategies following Poisson, uniform, normal and Chi-squared distributions. In [17] the authors studied the impact of different node distributions on energy consumption in WSN. It was shown that normal distribution outperforms uniform and expoential node distribution in terms of energy consumption and throughput. In this work the authors extend the work done in [17] by applying different node distributions in HE-WSN settings.

III. WSN Settings and Assumptions

The following models are assumed in this paper. The energy model is the same one used in [9].

A. SEP: Heterogeneous-Aware Clustering Protocol

Stable election protocol (SEP) [7] is a well-known clustering algorithm used specifically for HE-WSN. Heterogeneity in energy capabilities introduces the problem of speed death of the network. The basic operation of the SEP protocol is to utilize the heterogeneity in energy storage of sensor nodes to prolong the lifetime of the WSN. In SEP sensor nodes are divided into two categories: advanced and normal nodes. The advanced nodes represents a fraction (m) of the total number of nodes that are have more energy (α) than the other normal nodes. If a normal node has a basic energy storage of E_o then an advance node will have $(1+\alpha)E_o$. Then the total energy in the HE-WSN is given by:

$$\begin{aligned} E_{net} &= m*n*E_{advance} + (1-m)*n*E_{normal} \\ &= m*n*E_o*(1+\alpha) + (1-m)*n*E_o \\ &= n*E_o*(1+\alpha) \end{aligned} \quad (1)$$

SEP will utilize the advanced nodes as CHs more than the normal nodes in the same extended epoch compared to an epoch in HO-WSN. SEP will divide a protocol extended epoch into h sub-epochs. Normal nodes are elected once in a protocol extended epoch while advanced nodes are elected h times. In this way, network energy expenditure is distributed efficiently by the nodes and the network lifetime is extended.

B. Node Distributions Parameters Estimation

In this paper three different node distributions in a network area are utilized, namely, uniform, normal and exponential. The problem with distributions other than the uniform distribution is the distribution's parameters. Maximum likelihood estimation (MLE) is used to solve this problem. Node's locations are generated by the uniform distribution since all the information required by the distribution is provided, the information include the number of nodes and the dimensions of the network area. MLE will utilize the location information of the nodes provided by the uniform distribution and estimate the necessary parameters for other distributions. Table I shows the necceasy parameters required by the normal and exponential distributions. Paper [17] provides a complete derivation of these parameters. x_i represents a random variable of node position.

TABLE I: Estimated Distribution parameters

Distribution	Parameter 1	Parameter 2
Normal	$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$
Expoenential	$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i}$	–

C. Node Energy Communication Model

The main energy source of a sensor node is the battery. A node with more energy will live more compared to a node with less energy. HE-WSN lifetime will depend mostly on the nodes' energies and communication protocol used. A typical first order energy model used in HE-WSN is provided by Fig. 1 [17].

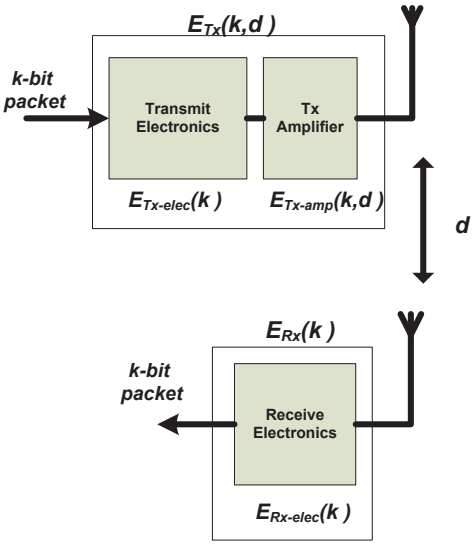


Fig. 1: Node communication model [17]

Energy expended by the transmitter to processes and sens k-bit packet is given by:

$$E_{TX}(k,d) = E_{TX-elec}(k) + E_{TX-amp}(k,d) \quad (2)$$

$$E_{TX-elec}(k) = k * E_{elec} \quad (3)$$

ISSN 1737-9334

$$E_{TX-amp}(k,d) = \begin{cases} k\epsilon_{fs}d^2, & if\ d < d_0 \\ k\epsilon_{mp}d^4, & if\ d \geq d_0 \end{cases} \quad (4)$$

where, E_{elec} is the amount of energy consumed by RF circuit in joules/bit for operating the receiver or transmitter circuits. ϵ_{fs} and ϵ_{mp} represent the free space and multiple path amplification models. d_0 represents the threshold distance by which a node can swap between the free space and multile path models. d_0 is computed using the formula $d_0 = \sqrt{\epsilon_{fs}/\epsilon_{mp}}$ [6].

On receiving a k-bit packet, a node will expend:

$$\begin{aligned} E_{RX}(k) &= E_{RX-elec}(k) \\ &= k * E_{elec} \end{aligned} \quad (5)$$

Local computation of s signals in a CH with k-bit length will consume:

$$E_{DA-tot} = s * k * E_{DA} \quad (6)$$

where E_{DA} represents the energy dissipated by aggregating one bit per signal.

IV. Performance Evaluation

In this section the performance measure metrics will be explained, then network settings and simulation results are analysed.

A. Performance Metrics

In this section the metrics used to evaluate the performance of HE-WSN with different node distributions are defined. These metrics are also used in [7].

- Stability and Instability Periods: Stability period or region represents the time interval until the death of the first node, while the instability period (unstable region) represents the time interval from the death of the first node to the death of the last node.
- Network lifetime: represents the time interval from the start operation of the WSN to the death of the last node.
- Network throughput: represents the total data rate reached the BS from the CHs.

B. Simulation Settings

Clustered HE-WSN is simulated in a region with dimensions $100m \times 100m$. Nodes are distributed according to three distributions, uniform, normal and exponential. The total number of nodes is 100. Nodes come into two flavors, normal and advanced. Uniform node placement strategy is used to estimate other distribution parameters utilizing MLE technique shown above. Table II shows the parameter settings used in the simulation.

TABLE II: Network parameters settings

Parameter	Value
Node energy	0.5 J
E_{elec}	50 nJ/bit
E_{DA}	5 pJ/bit
ϵ_{fs}	10 pJ/bit/m^2
ϵ_{mp}	0.0013 pJ/bit/m^4
k	4000 Bits
Number of nodes (N)	100
% of cluster heads (P)	0.05
Rounds	8k and 10k

C. Simulation Results

3 shows the number of alive nodes in LEACH based scenario for two cases. The first case is the plain case in which $m = 0$ and $\alpha = 0$, the second case is when $m = 0.2$ and $\alpha = 3$. It is known that LEACH protocol is used in HO-WSN and it is heterogeneous-oblivious protocol. In the first case, node distributions has no great effect on the stability region as the time it took for the first node to die (FND) is almost the same. In the second case LEACH takes advantage of the presence of advanced nodes as the FND for all node distributions increased significantly (longer stable region). The network lifetime (NLT) is also increased as the round of the last node to die (LND) is also increased.

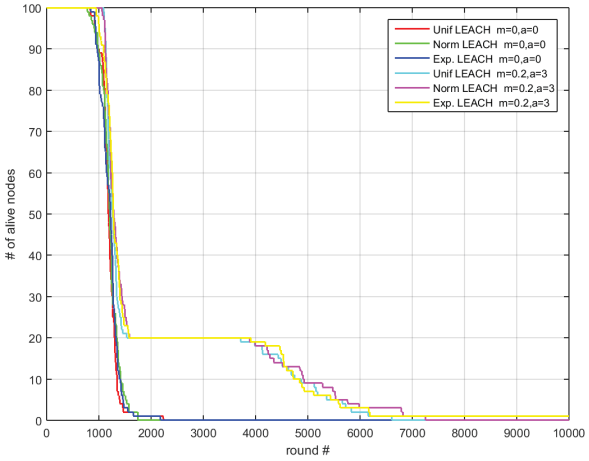


Fig. 2: No. of alive nodes using LEACH protocol

Figure 3 shows a comparison of SEP protocol for different node distributions. The simulation is done for two scenarios. In the first scenario the heterogeneity factors are $m = 0.2$ and $\alpha = 1$. In the second scenario the heterogeneity factors are $m = 0.2$ and $\alpha = 3$. It is noticed that the node distributions have small effect in extending the stability region of the HE-WSN. Uniform node distribution has the longest stable region compared to the other distributions. In the second scenario more energy is pumped into the advanced nodes, and as a result the stability region has increased for all node distributions with advantage of the uniform node distribution.

ISSN 1737-9334

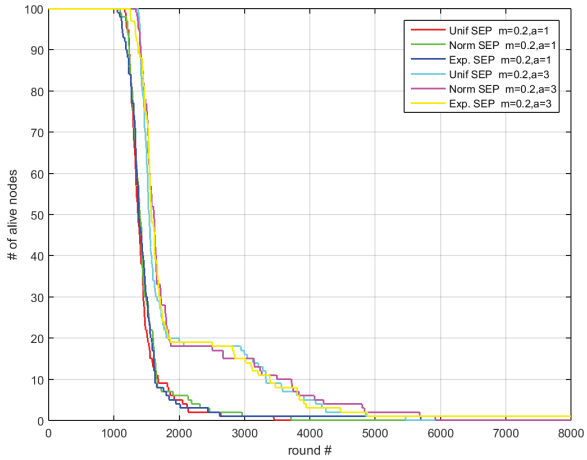


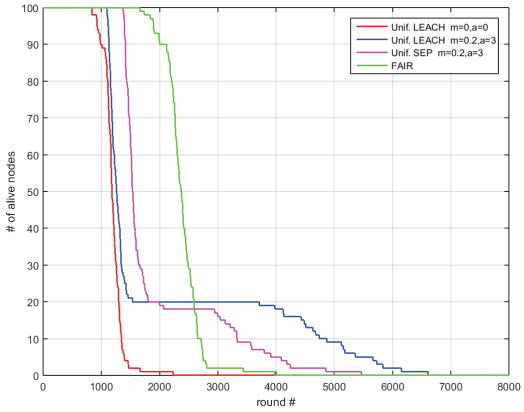
Fig. 3: No. of alive nodes using SEP protocol

Figure 4 compare the performance of SEP to: LEACH with the same heterogeneity factors m and α and LEACH where the extra energy of the advanced nodes is distributed uniformly over all WSN nodes, in this case it is refereed to as FAIR protocol. In all cases of different node distributions SEP take full advantage of the extra advanced nodes energies. The stable region in uniform and normal cases have increased by 25% and the exponential case by 30% compared to the LEACH with the same heterogeneity factors. The unstable region in SEP is shorter compared to the LEACH for all node distributions.

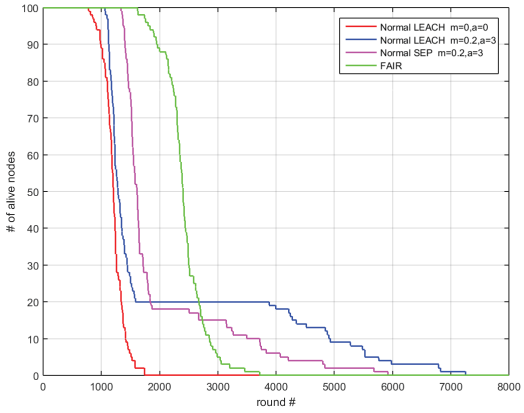
Figure 5 shows the network throughput received by the BS over the course of the network lifetime. It can be seen clearly that the SEP protocol throughput is significantly larger than LEACH in all cases of different node distributions as shown in sub-figures 5(a-c). This is because SEP will guarantee more CHs in more rounds compared to the LEACH protocol. This implies more data reports will be send to the BS. In sub-figure 5d the network throughput for different node distribution using SEP protocol is shown. It can be seen obviously that the normal distribution has the largest throughput over the course of the network lifetime. The reason for that is the distances of the nodes (normal and advanced) to the BS is smaller compared to the other distributions, which mean less energy is expended in data transmissions compared to the other distributions. The other reason is that in normal distribution CHs are always there during the rounds of the protocols which implies data packets are always reported to the BS compared to the other distributions.

V. Conclusion

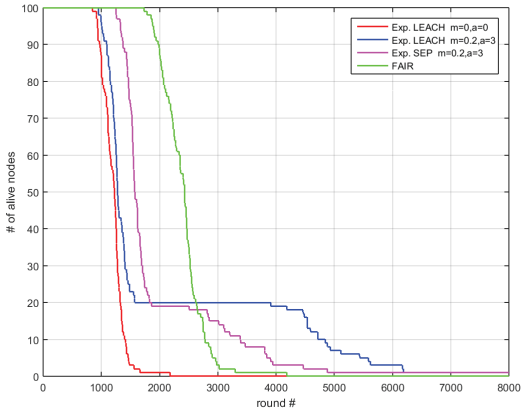
In this paper we studied the impact of different node distributions on the lifetime of heterogeneous WSN. Three different node distributions are utilized namely, uniform, normal and exponential. The parameters of the normal and exponential distributions are computed using the maximum likelihood estimation principle. SEP and LEACH protocols are utilized



(a) Alive nodes per round - uniform dist.



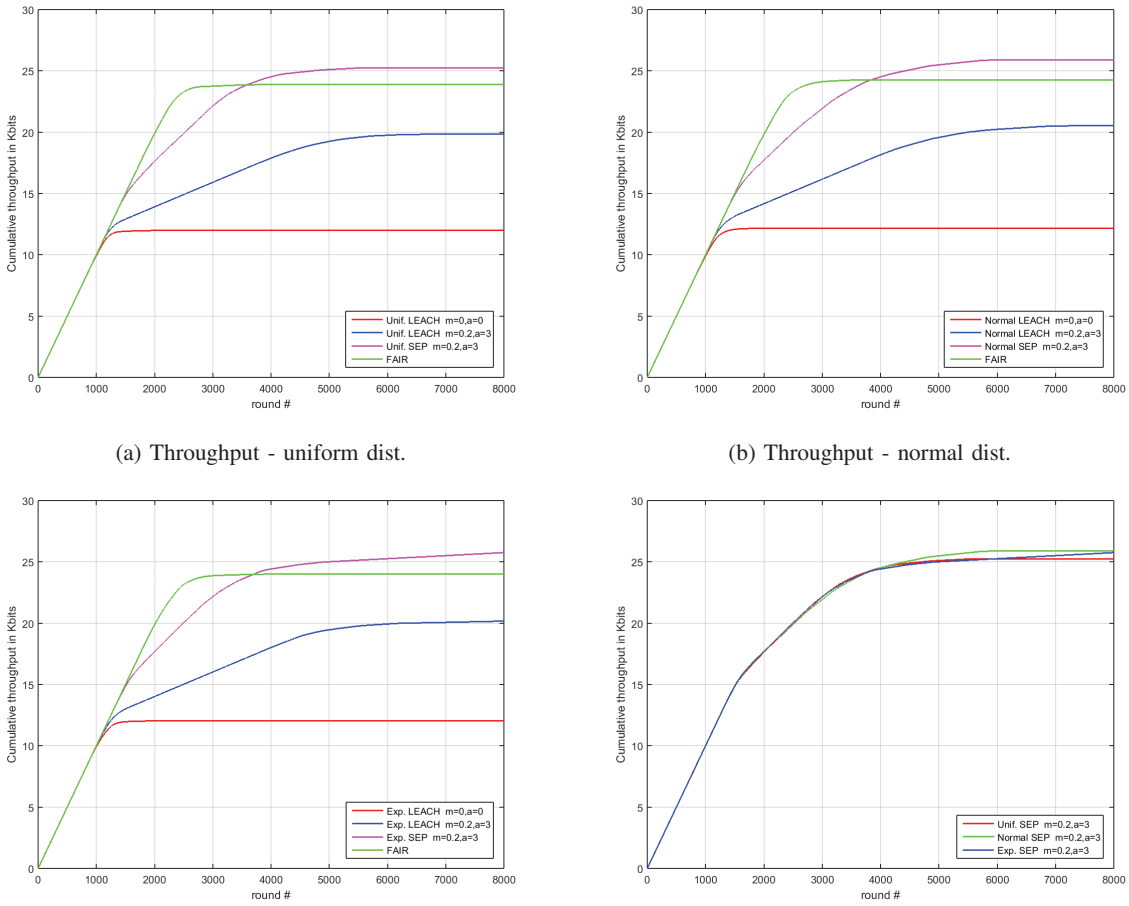
(b) Alive nodes per round - normal dist.



(c) Alive nodes per round - exponential dist.

Fig. 4: Different Node distributions

ISSN 1737-9334



(a) Throughput - uniform dist.

(b) Throughput - normal dist.

(c) Throughput - exponential dist.

(d) Throughput - All distributions.

Fig. 5: Network throughput

in this study. SEP protocol is used in HE-WSN while LEACH protocol is used mainly in HO-WSN. Stability and instability periods a long with the network lifetime are utilized as metrics for evaluating the performance of SEP and LEACH in HE-WSN. It was noticed that node distributions have small effect of the stability regions of heterogeneous-aware (SEP) protocol and heterogeneous-oblivious (LEACH) protocol. In all cases of different node distributions SEP protocol outperforms LEACH with the same heterogeneity factors in stability region and unstable region metrics. In throughput analysis normal distribution shows slightly greater data rate compared to the uniform and exponential distributions.

ACKNOWLEDGMENT

This research is under a project titled “Load Balancing Algorithm for Green Computer Networks”. The project is funded by Al-Zaytoonah University of Jordan. Resolution number 18/64/2016-2017.

REFERENCES

[1] H. Barani, Y. Jaradat, H. Huang, Z. Li, and S. Misra. ”Effect of sink location and redundancy on multi-sink wireless sensor networks: a capacity and delay analysis.” IET Communications 12, no. 8 (2018): 941-947.

[2] Y. Wang, P. Shi, K. Li, and Z. Chen. ”An energy efficient medium access control protocol for target tracking based on dynamic convey tree collaboration in wireless sensor networks.” International Journal of Communication Systems 25, no. 9 (2012): 1139-1159.

[3] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. ”A survey on sensor networks.” IEEE communications magazine 40, no. 8 (2002): 102-114.

[4] D. Bhattacharyya, T. Kim, and S. Pal. ”A comparative study of wireless sensor networks and their routing protocols.” Sensors 10, no. 12 (2010): 10506-10523.

[5] Singh, Sunil Kumar, Prabhat Kumar, and Jyoti Prakash Singh. ”A survey on successors of LEACH protocol.” IEEE Access 5 (2017): 4298-4328.

[6] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan. ”Energy-efficient communication protocol for wireless microsensor networks.” In System sciences, 2000. Proceedings of the 33rd annual Hawaii international conference on, pp. 10-pp. IEEE, 2000.

[7] G. Smaragdakis, I. Matta, and A. Bestavros. SEP: A stable election protocol for clustered heterogeneous wireless sensor networks. Boston University Computer Science Department, 2004.

[8] Y. Jaradat, M. Masoud, I. Jannoud, T. Abu-Sharar, and A. Zerek. ”Performance Analysis of Homogeneous LEACH Protocol in Realistic Noisy WSN.” In 2019 19th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), pp. 590-594. IEEE, 2019.

[9] Y. Jaradat, M. Masoud, and I. Jannoud. ”A Mathematical Framework

ISSN 1737-9334

of Optimal Number of Clusters in 3D Noise-prone WSN Environment." IEEE Sensors Journal (2018).

[10] H. Barani, Y. Jaradat, H. Huang, Z. Li, and S. Misra. "The effect of popularity rule on capacity and delay in multi-sink WSNs." In 2017 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1-6. IEEE, 2017.

[11] Y. Chen, C. Chuah, and Q. Zhao, "Sensor placement for maximizing lifetime per unit cost in wireless sensor networks", IEEE Military Communications Conference, MILCOM 2005, October 2005, pp. 1097-1102

[12] G. Deepak and B. Beferull-Lozano, "Power-efficient Sensor Placement and Transmission Structure for Data Gathering under Distortion Constraints", ACM Transactions on Sensor Networks (TOSN), Vol. 2, Issue 2, May 2006, pp. 155 – 181.

[13] S. H. Al-Sharaeh, et.al, "Multi-Dimensional Poisson Distribution Heuristic for Maximum Lifetime Routing in Wireless Sensor Network", World Applied Sciences Journal 5 (2): 119-131, 2008 ISSN 1818- 4952.

[14] W. F. Yun Wang and D. P. Agrawal, " Gaussian versus Uniform Distribution for Intrusion Detection in Wireless Sensor Networks", IEEE Transactions on Parallel And Distributed Systems, Vol. 24, No. 2, February 2013.

[15] A. E. Omar Said, "Optimizing Sensors Distribution for Enhancing WSN Intrusion Detection Probability in Euclidian's Space", International Journal of Computer Applications (0975 – 8887) Volume 105 – No. 15, November 2014.

[16] F. M. Osman and S. H. Al-Sharaeh," Hetrogeneous Multi-Deployment Strategy Effect on Maximizing the Lifetime Routing in Wireless Sensor Network", Middle-East Journal of Scientific Research 13 (6): 749-759, 2013 ISSN 1990-9233

[17] Y. Jaradat, M. Masoud, I. Jannoud, and D. Zaidan. "The Impact of Nodes Distribution on Energy Consumption in WSN." In 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), pp. 590-595. IEEE, 2019.

ISSN 1737-9334

Fractional order PID Controller with Online tuning for a Class of Fractional Order Plants based on Bode's Ideal Transfer Function, FRIT and RLS

A.Dif[#], A.Charef[*] and A.Djouambi[#]

[#] *Laboratoire de Génie Electrique et Automatique (LGEA)- Département des Sciences et Sciences Appliquées Université d'Oum-El-Bouaghi 04000. Algérie.*
ahlem_dif@yahoo.fr
djouambi_abdelbaki@yahoo.fr

[*] *Département d'Electronique- Université Constantine 1 Route Ain El-bey Constantine 25011, Algeria.*
afcharef@yahoo.com

***Abstract*—Fractional order dynamic model could model various real materials more adequately than integer order ones and provide a more adequate description of many actual dynamical processes. Fractional order controller is naturally suitable for these fractional order models. In this paper, a new strategy for digital control and parameters identification of robust fractional order controllers applied to a class of fractional order system models is presented. The proposed strategy is based on fractional reference model. The model consists on an ideal closed-loop system whose open-loop is given by the Bode's ideal transfer function with the suitable parameters. This technique suggests an online parameters tuning of fractional order $PI^{\alpha}D^{\beta}$ controller (FPID) using the Fictitious Reference Iterative Tuning method (FRIT). After having set the optimal non-integer orders α and β, the gains of the FPID are then estimated in a recursive manner using the so-called Recursive Least Squares (RLS) algorithm with forgetting factor, which can cope with variation of plant characteristics adaptively.**
Better performance using fractional order $PI^{\alpha}D^{\beta}$ controllers can be achieved and is demonstrated through an example with a comparison to the classical integer order PID controllers for controlling fractional order systems (FOS).

Keywords—fractional order system; fractional order PID control; fictitious reference iterative tuning; recursive least square algorithm; online tuning; fractional reference model; Bode's ideal loop.

I. INTRODUCTION

Most theoretical control system design methods are based on mathematical models. However, it is often difficult to collect input-output data for the identification of a controlled plant that has already been set up and is in full operation. Therefore direct approaches based on input-output measurements with no need for help from a plant model have attracted attention from several researchers.

Among the direct controller parameter tuning methods, fictitious reference iterative tuning (FRIT) proposed in [1] which is performed on input and output data obtained from only a one-shot experiment. FRIT based PID control (FRIT-PID) can be considered as a free model control approach, it has received much attention recently as a practical and useful method, and its extended methods have been studied in [2]; [3]; [4]. Therefore, this method can be implemented online where the controller parameters have to be updated periodically [3]; [5].

Given the large number of discoveries on the fractal dimensions of nature and science/technology fields, fractional calculus and its corresponding applications have attracted considerable research interest, and many excellent results have been obtained in many scientific areas of physics and engineering [6]; [7]; [8]; [9]. Fractional (or non integer order) systems are considered as a generalization of integer order systems [10], [11]. The advantages of fractional form become apparent in modeling mechanical and electrical properties of real devices as well as in the theory of dynamical systems. Sometimes, mathematical modeling of processes and physical phenomena leads to differential equations of fractional orders and so fractional order calculus is more necessary and accurate than the integer order calculus to describe or explain their real dynamic behavior.

For example, Reyes-Melo et al. [8] indicated that the dielectric relaxation phenomena in polymeric materials can be clearly described by using a fractional order model. Meral et al. [13] confirmed that the essential properties of viscoelastic materials can be precisely reflected by using a fractional order model.
Le Méhauté and Crepy [37] have proposed a concept of a fractance—a new electrical circuit element, which has intermediate properties between resistance and capacitance. Recent research on the modeling of real inductors and capacitors shows that the nature of these two circuit elements are fractional orders and that these elements should be modeled by using fractional calculus [38]; [9]. Furthermore, Boulgameh et al [39] have validated experimentally the dynamical fractional order model of PN junction diode based small signal equivalent circuits.

The interest of these fractional order operators has been motivated by their good performances obtained in system identification and control theory; from where their introduction in the synthesis of PID control systems has allowed having some remarkable results in term of robustness quality compared with the conventional PID controller.

Oustaloup in his book [27] has developed a new strategy of control named CRONE, which is the acronym in French of 'Commande Robuste d'Ordre Non Entier', i.e., robust control of non integer order or robust control of fractional order. He has applied it in various fields of systems control [30]. More recently, Podlubny [12] has proposed a generalization of the PID controller, namely the fractional order controller $PI^{\alpha}D^{\beta}$, involving non integer actions α and β of integration and derivation respectively.

Several methods have been proposed for the synthesis and the control design PID. These strategies are known as reference model tuning and consist to obtain a control guaranteeing that the transfer function of the controlled system, with unit

ISSN 1737-9334

feedback is equivalent to Bode's ideal transfer function. Using a closed loop system with Bode's ideal transfer function as reference model makes closed loop controlled system robust to gain variations and step responses exhibiting an iso-damping property [28]; [31]; [32]. Furthermore, it was proved in many works, e.g. [29]; [33]; [34]; [35], the ability to design the fractional controlled system using Bode's ideal transfer function as a fractional reference model.

In order to demonstrate the necessity of a fractional order controller $PI^{\alpha}D^{\beta}$ for a more efficient control of fractional order systems, we propose in this paper, an extension of the standard FRIT-PID to fractional order control using fractional order controller $PI^{\alpha}D^{\beta}$, using the Bode's ideal loop, denoted (BIL), of the FRIT-PID approach as a fractional reference model, thinks to its advantageous features cited above, instead of integer order model. So, in this context, we have designed a FRIT approach with a fractional order system (FS) - fractional order reference model (BIL) and fractional order control.

The parameters tuning of the $PI^{\alpha}D^{\beta}$ is achieved online using the so-called recursive least squares (RLS) algorithm with forgetting factor [14], the numerical evaluation of the fractional operators is carried out using the DAFI filter proposed in [15].The suggested approach, denoted FRIT-FPID, will be compared with the FRIT-PID method proposed in [16] of course with BIL as a fractional reference model.

This paper is divided in five parts. The first one presents some mathematical materials used to approximate and simulate fractional order systems. The second part is devoted to the presentation of Bode's ideal transfer function which will be proposed as a fractional reference model with the suitable parameters. Part three presents the FRIT-PID method. In the fourth part, an online tuning of controller parameters via RLS algorithm is presented. Part five, establishes the proposed FRIT-FPID approach and its online implementation using DAFI filter and RLS algorithm with forgetting factor. An example is provided to demonstrate the necessity of such controllers for the more efficient control of fractional order systems. Finally, we draw the main conclusions and address perspectives towards future developments.

II. FRACTIONAL CALCULUS AND FRACTIONAL ORDER DYNAMIC SYSTEMS

The idea of fractional calculus has been known since the development of the regular (integer-order) calculus, with the first reference probably being associated with Leibniz and L'Hôpital in 1695 where half-order derivative was mentioned [40].

A. Definitions

Fractional calculus is a generalization of integration and differentiation to fractional order operator denoted by ${}_aD_t^{\alpha}$, where a is related to the initial conditions and $(\alpha \in R)$ is the operation order. They are represented by the following irrational transfer function:

$$G(s) = \frac{\mathrm{Y}_{\alpha}(\mathrm{s})}{\mathrm{Y}(\mathrm{s})} = \frac{1}{s^{\alpha}} \tag{1}$$

where s = jw is the complex frequency and $\alpha \in R$.

There are several definitions of fractional integration and differentiation [41]. A commonly used definition is the Riemann-Liouville (RL) given by [24]:

$${}_aD_t^{\alpha} f(t) = \frac{1}{\Gamma(m-\alpha)}\left(\frac{d}{dt}\right)^m \int_a^t \frac{f(\tau)}{(t-\tau)^{1-(m-\alpha)}} d\tau \tag{2}$$

where m is an integer number, $m-1 < \alpha < m$ and $\Gamma(\cdot)$ is the well-known Euler's gamma function.

An alternative definition, based on the concept of fractional differentiation, is the Grünwald-Letnikov (GL) definition given by:

$${}_aD_t^{\alpha}\mathrm{y(t)} = \lim_{\mathrm{h}\to 0} \frac{1}{h^{\propto}} \sum_{k=0}^{[(t-a)/h]} (-1)^k \binom{\alpha}{k} y(t-kh) \tag{3}$$

where $[\cdot]$ denotes the integer part and h is the sampling time. The Laplace transform of the fractional operator, under zero initial conditions for real order α is given by [41]:

$$L\left\{ {}_aD_t^{\alpha}\mathrm{y(t)}\right\} = s^{\alpha} Y(s) \tag{4}$$

B. Fractional order system and their simulation

Fractional order system is a dynamical system that can be modeled by fractional differential equation using Fractional order differentiator [42]. A major difficulty with fractional models is its time-domain simulation. Often, the analytical solution of fractional differential equation is not simple to compute. During the last 20 years, many methods are developed in order to approximate and simulate this type of irrational models. Two types of methods can be considered. The first ones, also called indirect methods, are based on the simulation of the continuous fractional model with the help of a specific operator or representation [17]; [18]; [19]; [20]; [21], [22]. In [23] an approximation of fractional order operators by rational functions is presented; the differentiator or integrator is approximated in a given frequency band $[\omega_l, \omega_h]$ by a rational function using a set of elementary first order cells.

The second ones, called direct methods, are based on numerical approximation of the fractional order operator. The most commonly used is that directly related to GL definition given by:

$$\frac{{}_a d^{\alpha}_t y(t)}{dt^{\alpha}} \approx \frac{1}{h^{\propto}} \sum_{k=0}^{N} (-1)^k \binom{\alpha}{k} y(t-kh) \tag{5}$$

where *N* is a limited number of terms, fixed according to the short memory principle [25].

Grünwald-Letnikov approximation is very simple to use. However, the simulation requires, for each step, the computation of sums of increasing dimension with time. This makes real-time simulation hard to achieve and amplify greatly noise present in the data. This is a real constraint when this approximation is used for model identification and parameter estimation. To cope with this problem, a new technique for numerical simulation of fractional order systems in a given frequency band of interest $[\omega_L, \omega_H]$ is proposed in [15] using the so called Digital Adjustable Fractional Integrator (DAFI). Fig. 1 presents the block diagram of this filter used for numerical simulation of fractional operators given in equation (1).

where h_i , (i=1,2,..,N), are given by:

$$h_i(\alpha) = \left[\omega_c 10^{\varepsilon(0.2\alpha-0.1)}\right]^{-\alpha} \frac{\prod_{j=0}^{N-1}\left(1-10^{0.4\varepsilon(i-j-\alpha)}\right)}{\prod_{j=0,j\neq i}^{N-1}\left(1-10^{0.4\varepsilon(i-j)}\right)} \tag{6}$$

such that $\omega_c \ll \omega_L$, ε and N represent the approximation error and the filter dimension respectively.

ISSN 1737-9334

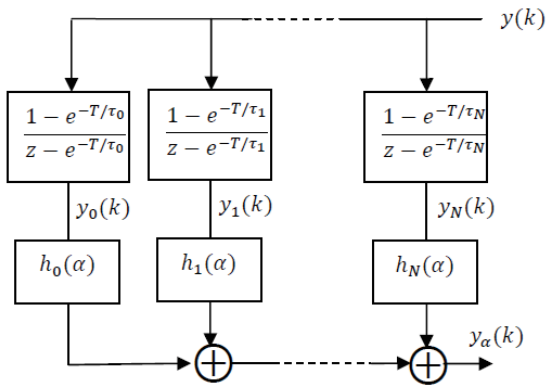


Fig. 1. Digital adjustable fractional order integrator (DAFI)

$$N = integer\left[\frac{log\left(\frac{\omega max}{p_0}\right)}{log\,(ab)}\right] + 1 \tag{7}$$

$$p_0 = \omega_c 10^{\left(\frac{\varepsilon}{20\alpha}\right)},\ ab = 10^{\left(\frac{\varepsilon}{10\alpha(1-\alpha)}\right)} \tag{8}$$

where

$\omega_{max} \gg \omega_H$. The output y_α is then given by:

$$y_\alpha(k) = \sum_{i=0}^{N} h_i(\alpha) y_i(k) \tag{9}$$

and

$$y_i(k) = e^{-\frac{T}{\tau_i}} y_i(k-1) + \left(1 - e^{-\frac{T}{\tau_i}}\right) y(k-1) \tag{10}$$

$$\tau_i = \frac{1}{p_i(0.5)} \tag{11}$$

where T is the sampling time.

III. Bode's Ideal Transfer Function as a Reference Model

In 1945, H.W. Bode has presented in his book a design of feedback amplifiers [36]. He has proposed an ideal shape of the open loop transfer function of the form:

$$M(s) = \left(\frac{\omega_u}{s}\right)^{\mu}, \quad 1 < \mu < 2 \tag{12}$$

where

ω_u is the gain crossover frequency, that is $|M(j\omega_u)| = 1$.

The open-loop Bode diagrams Fig. 2 have an amplitude slope of -20μ dB/dec and a constant phase of $-\mu\pi/2$ rad. Therefore, the closed-loop system Fig.3 has a constant phase margin of $PM = \pi(1-\mu/2)\,rad$, that is independent of the system gain $K_u = \omega_u^{\mu}$. This makes closed-loop systems robust to gain variations and, therefore, the closed loop system with a unit feedback may be used as a typical reference model in order to synthesize robust control systems.

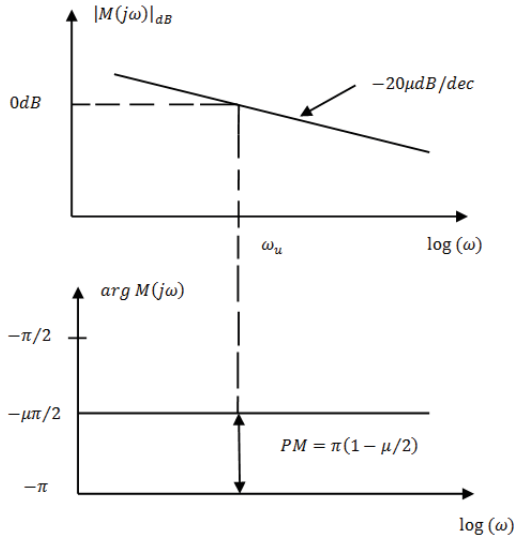


Fig. 2. Bode diagrams of amplitude and phase of $M(s)$ for $1 < \mu < 2$.

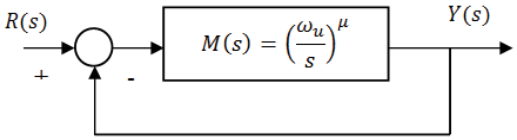


Fig. 3. Fractional-order control system with Bode's ideal transfer function $M(s)$.

The desired specifications of the step response of the reference system can be achieved using the appropriate values of the parameters μ and ω_u [28].

IV. The FRIT Method

Consider the unity-feedback control system shown in Fig. 4, where the process is a stable or unstable. In the figure, $C(\theta, s)$ is the controller, and $\boldsymbol{\theta}$ denotes a parameter vector to be tuned in the controller. Also, u, y, r, and $e = r - y$ denote the control input, output, reference signal, and error signal, respectively.

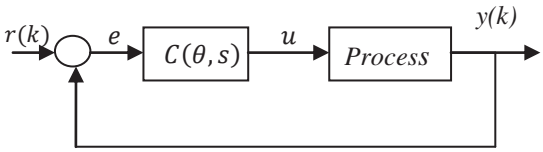


Fig. 4 The unity-feedback control system.

Let us consider the case where the feedback controller $C(\theta, s)$ is a fractional order PID (FPID) linearly parameterized in terms of adjustable control parameters vector θ. That is, $C(\theta, s)$ can be expressed as:

$$C(\theta, s) = \theta^T \varphi(s) \tag{13}$$

where

$$\theta = \left[k_p, k_i, k_d\right]^T \tag{14}$$

is the controller parameters and

$$\varphi(s) = [1,\ s^{-1},\ s]^T \tag{15}$$

is the PID controller operators.

ISSN 1737-9334

FRIT is a direct control parameter tuning method using one-shot experimental input-output data with no need for help from a plant model. FRIT searches PID control parameters so that the plant output follows the reference model output, which specifies an ideal response for the reference signal. However, if an inappropriate reference model is selected, FRIT might possibly derive improper control parameters [26].

The tuning method minimizes the following performance index [3]:

$$\hat{J}(\theta) = \sum_{k=1}^{N}[\hat{e}(k)]^2 \tag{16}$$

where

$$\hat{e}(k) = \theta^T \xi(k) - d(k) \tag{17}$$

$$\xi(k) = \emptyset(z)\big(1 - M(z)\big)y(k) \tag{18}$$

$$d(k) = M(z)u(k) \tag{19}$$

where $M(z)$ is the reference model.

The minimization problem of $\hat{J}(\theta)$ is regarded as a least-squares problem because $\hat{e}(k)$ is linear with respect to θ.

V. Online Tuning via Recursive Least-Squares Method

In [3], a normalized recursive least square method for the adaptive adjusting law for the model reference adaptive control based on an online FRIT approach is proposed. The boundedness of all signals in the closed loop system as well as asymptotically tracking the reference model output was proved.

Although this tuning procedure is carried out online, the controller parameters can be abruptly updated at a definite period of time, so that the control performance may be deteriorated. Moreover, the computational complexity of solving the normal equation is relatively large.

To resolve these problems, the authors in [16] utilize the so called RLS method with forgetting factor [14] and propose a filtering rule for reducing the variations of controller parameters.

The RLS algorithm with a forgetting factor λ, $(0 < \lambda < 1)$, is then given by:

$$h(k) = P(k-1)\xi(k)/\big(\lambda + \xi(k)^T P(k-1)\xi(k)\big) \tag{20}$$

$$P(k) = \Big(P(k-1) - h(k)\big(\xi(k)^T P(k-1)\big)\Big)/\lambda \tag{21}$$

$$\hat{\theta}(k) = \hat{\theta}(k-1) + h(k)\left(d(k) - \xi(k)^T \hat{\theta}(k-1)\right) \tag{22}$$

where $\hat{\theta}(0)$ is the initial controller parameter and $P(0)$ is the initial correlation matrix usually set to be:

$$P(0)=\gamma I,$$

where $\gamma > 0$ is a large constant for high signal to- noise ratio.

The filtering rule of the implemented controller parameters is given by:

$$\theta(k) = (1-m)\theta(k-1) + m\,\hat{\theta}(k-1) \tag{23}$$

where m is a sufficiently small positive constant.

VI. The Proposed Online Fractional Order Controller Parameter Tuning Algorithm Using The Bode's Ideal Transfer Function

In this section a generalization of the FRIT method to fractional order is proposed. The classical controller PID is generalized to a fractional PID denoted FPID in order to control the fractional order system. The proposed method, denoted FRIT-FPID serves to tune the FPID parameters using Bode's ideal loop BIL as a fractional reference model.

The proposed tuning strategy starts by the determination of the parameter μ of Bode's open loop transfer function according to the desired control performances such that the phase margin $\emptyset_m$ and the unity gain frequency ω_u in a given frequency band $[\omega_L, \omega_H]$ around ω_u. The suggested approach is implemented online using DAFI filter and RLS algorithm with forgetting factor. The transfer function of the fractional order controller can be given as [12]:

$$C(s) = k_p + k_i \frac{1}{s^\alpha} + k_d s^\beta \ , \quad \alpha, \beta \in R \tag{24}$$

where k_p, k_i and k_d are the proportional, integral and derivative gains, respectively, and α and β are the fractional integrator and fractional derivator respectively.

The regressive form of the fractional order controller can be given by:

$$C_c(s,\theta) = \theta^T \emptyset_{cf}(s) \tag{25}$$

where

$$\theta = \left[k_p, k_i, k_d\right]^T \tag{26}$$

$$\emptyset_{cf}(s) = \left[1, \frac{1}{s^\alpha}, s^\beta\right]^T \tag{27}$$

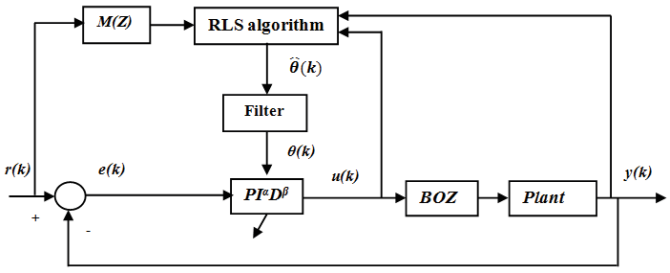


Fig. 5. Block diagram of online Fractional control system tuning

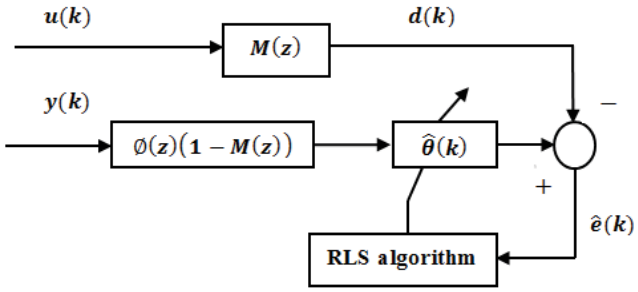


Fig. 6. Block diagram of RLS algorithm

Fig. 5, shows the block diagram of the online fractional control system tuning. Moreover, a more detailed mechanism of the RLS algorithm is illustrated in Fig. 6.

The design of the fractional order integrator α and the fractional differentiator β of the FPID controller has started by using offline experiments. We first running the tuning algorithm in a limited time interval $[0, K]$, with a given arbitrary initial parameters $\theta_0(0)$, α_0 and β_0, where

ISSN 1737-9334

$0 < \alpha_0, \beta_0 < 1$. In the next step, we running the algorithm for several chosen values α_i and β_i, $0 < \alpha_i, \beta_i < 1$, where the initial vector parameters $\theta_i(0)$ of the i-th experiment is the final tuned parameters $\theta_{i-1}(K)$ of the previous one. The optimal values of α and β are then given by:

$$(\alpha, \beta)_{\text{op}} = \underset{0<\alpha_i,\beta_i<1}{\operatorname{argmin}} \sum_{k=1}^{K} \big(y_p(k) - y_m(k)\big)^2 \qquad (28)$$

where $y_p(k)$ is the controlled system output and $y_m(k)$ is the model output. Controller parameters tuning, with the designed optimal orders $(\alpha, \beta)_{\text{op}}$, can then be performed online to cope with variation of plant characteristics and disturbances adaptively.

VII. ILLUSTRATIVE EXAMPLE

In this section we give an example showing the usefulness of the proposed method FRIT-FPID in comparison with the conventional method FRIT-PID, applied to a fractional order system which plays the role of «reality», and its integer order approximation which plays the role of «model».

A. *Fractional Order Controlled System*

We consider the unity feedback control fractional-order system (Fig. 4), which plays the role of "reality" given by [43] as:

$$G_p(s) = \frac{1}{a_2 s^{\beta 2} + a_1 s^{\beta 1} + a_0} \qquad (29)$$

Where we take a_2 =0.8, $a_1 = 0.5, a_0 = 1$, $\beta_2 = 2.2, \beta_1 = 0.9$.

with the following specifications:

· Phase margin = 35°;

• The settling time Ts = 1.1 s, corresponding to the unity gain crossover frequency of ω_u= $10\ rad/s$.

The open loop Bode's ideal transfer function assuring these specifications in a frequency range $[w_L, w_H]$ = $[0.1\omega_u, 10\omega_u]$ = [1 rad/s, 100 rad/s] around ω_u , must be chosen as:

$$G(s) = \frac{1}{(s/\omega_u)^{\mu}} = \frac{1}{(s/10)^{1.61}} \qquad (30)$$

The reference signal $r(k)$ is given by a unit step with an amplitude of 1.

According to the desired output $y_d(k) = M(z)r(k)$.

The sampling time is chosen to be 0.01. We apply the RLS algorithm and set the initial controller Parameters as:

$\theta(0) = [1{,}40{,}55]^T$, we set the parameters in the algorithm as: $\gamma = 10^2$, $m = 10^{-2}$ and $\lambda = 1 - 5.10^{-5}$.

Fig. 7, 8 and 9 show the gains evolution and the output responses using FRIT-PID and the proposed FRIT-FPID methods. We can clearly see that the tuned PID and FPID parameters converge to the optimal values rapidly in both methods and the best tracking performances is achieved using the proposed FRIT-FPID method.

Fig. 9 shows that the fractional order controlled system has well followed the fractional order reference model (BIL) using the proposed FRIT-FPID method, contrary to the response using the classical FRIT-PID method, where fractional order controlled system cannot follow the (BIL).

The fractional order system stabilizes rapid at $t = 1{,}5s$ and has a limited surplus oscillations comparatively to results obtained in [43], where the fractional order system has stabilized at $t \approx 5s$.

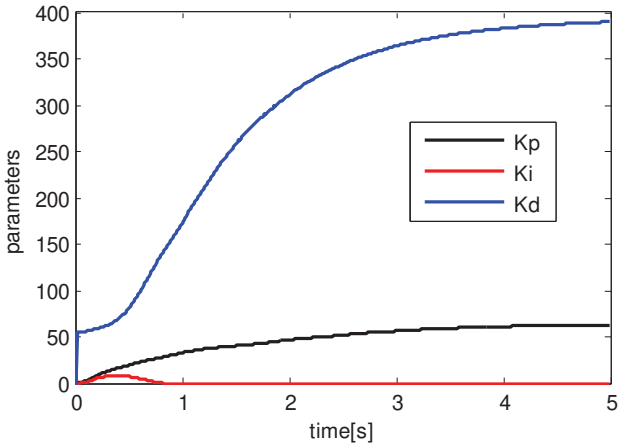


Fig. 7. Evolution of PID parameters.

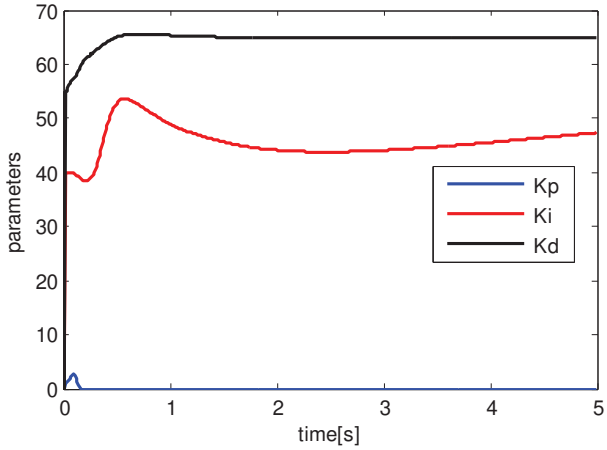


Fig. 8. Evolution of FPID parameters.

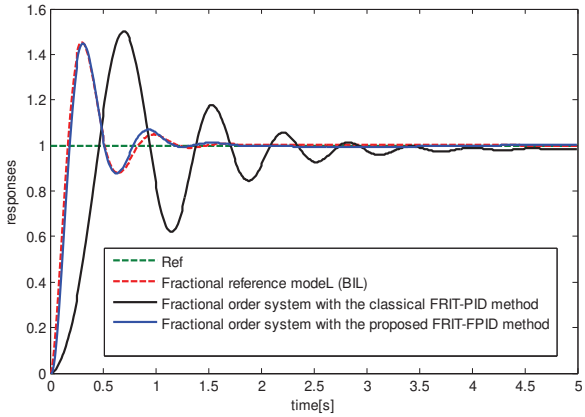


Fig. 9. Output using the conventional FRIT-PID method and the proposed FRIT-FPID method ($\alpha = 0.9$, $\beta = 0.9$).

The best results obtained by the proposed method can be justified by the flexibility of the fractional $PI^{\alpha}D^{\beta}$ controller that has two more degrees of freedom α and β than the classical one. This last has a fixed structure which limits the overall possible loop shapes of the fractional reference model.

The classical PID controller is shown to be not so suitable for the control of the fractional order system.

B. *Integer Order Approximation*

For the comparison purpose, let us approximate the considered fractional-order system by a second-order system. (Noticing that $\beta_2 = 2.2$, $\beta_1 = 0.9$ are close to 2 and 1 respectively).

ISSN 1737-9334

Fig. 10 and 11 show the gains evolution and the output response using FRIT-PID for controlling the integer order system (IS).

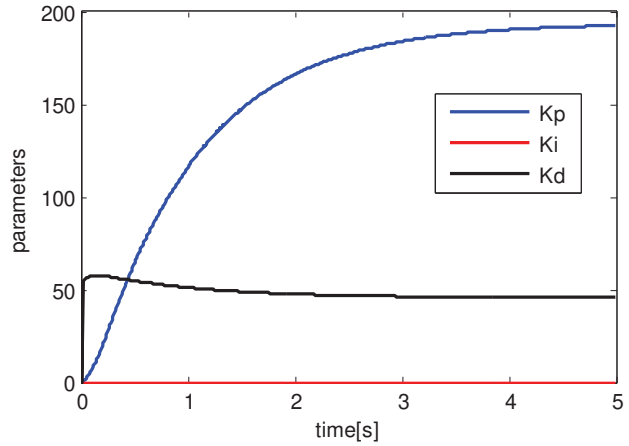


Fig. 10. Evolution of PID parameters (for an integer order system: IS).

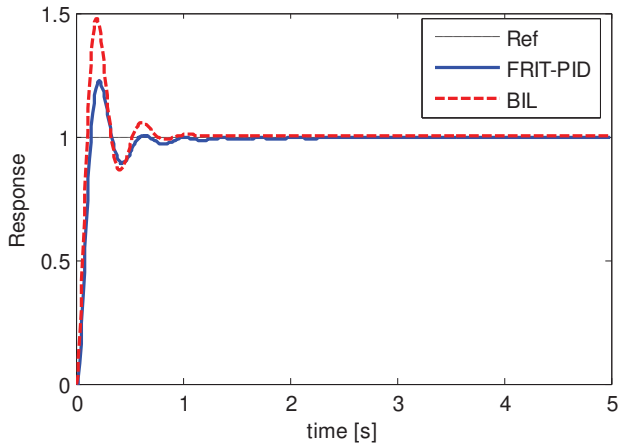


Fig. 11. Output using FRIT-PID method (IS IC.FM)

It's clear from Fig. 11 that the integer order controlled system converges to the reference signal, but it cannot follow the fractional reference model (BIL). We explain this by the fixed structure of the classical PID controller which limits the overall possible loop shapes of the fractional reference model.

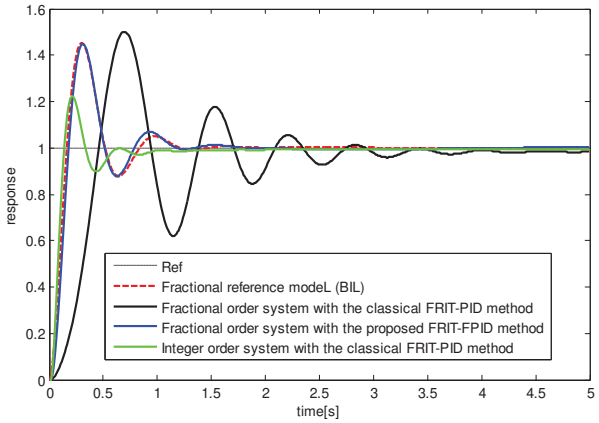


Fig. 12. Comparison of the unit-step response of the closed-loop (FS) and (IS) using the conventional FRIT-PID and the proposed FRIT-FPID methods.

Fig. 12 gives a comparison of the unit-step response of the closed loop with the fractional-order system controlled by fractional order $PI^{\alpha}D^{\beta}$ controller, the unit-step response of the closed loop with the same fractional-order system controlled by the integer-order PID controller and the closed loop with the integer order system controlled by a classical order *PID* controller.

The comparison shows that for satisfactory feedback control of the fractional order system using BIL as a fractional reference model it is recommended to use a controller of the similar nature; i.e., a fractional order $PI^{\alpha}D^{\beta}$ controller instead of a classical *PID* controller.

In order to evaluate the robustness of the proposed method in the case of fractional order system, which plays the role of «reality» compared with the integer order approximation, which plays the role of « model », we perform the following tests:

A. Abrupt Plant parametric variation test:

Consider the case where the gain of the fractional order plant is abruptly increased with 5% at t=2 s, i.e., the plant is 1,05*G_p after 2 s.
Fig. 13 shows the output obtained by the proposed FRIT-FPID approach.

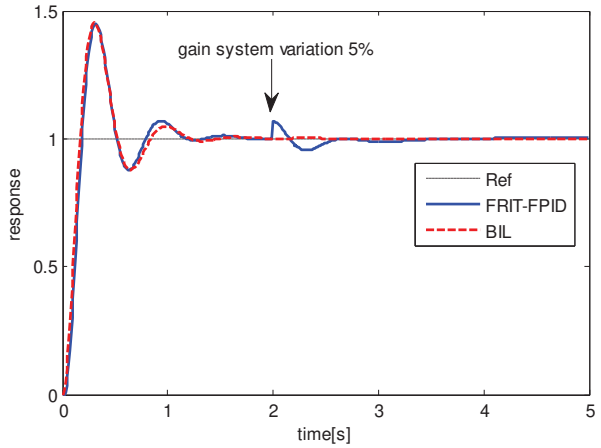


Fig. 13. Response of the fractional order system response in the case of parametric variation using the proposed FRIT-FPID, ($\alpha = 0.9, \beta = 0.9$)

After the fractional order plant parametric variation at time 2 s, the fractional order control system has well established this change. These results can be explained by the advantage of introducing the BIL as a fractional reference model, because it makes the fractional order system G_p robust to gain variations.

B. Output test with a reference noise:

We provide white noise with zero mean and standard deviation of 10^{-2} as reference disturbance. The output result using the proposed method is shown in Fig. 14.

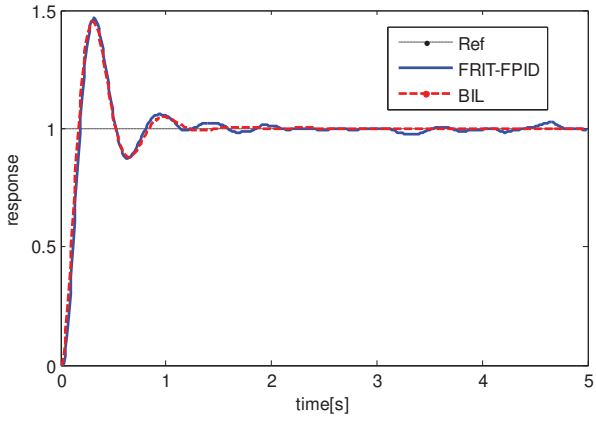


Fig. 14. Response with reference noise using FRIT-FPID method

($\alpha = 0.9,\ \beta = 0.9$)

ISSN 1737-9334

It is clear from the Fig. 14 that the output using the proposed FRIT-FPID method is less affected by the noise. This is due to the introduction of an online fractional $PI^{\alpha}D^{\beta}$ controller, that has two more degrees of freedom α and β than the classical one, which give a better flexibility to the control system to behave with this kind of disturbance.

Both tests show that the proposed FRIT-FPID technique is robust against fractional order plant parameters variation and measurement noise.

VIII. CONCLUSION

In this paper, we have provided a new and powerful design of the fractional order PID controller for the control of the fractional order systems, which play the role of reality. The proposed concept aims to tune the fractional order PID controller online using DAFI filter and RLS algorithm with forgetting factor.

The basic idea of the proposed technique, denoted FRIT-FPID, consists of the combination between the classical FRIT considered as a free model control approach, and a fractional order controller $PI^{\alpha}D^{\beta}$. The proposed approach uses Bode's ideal control loop (BIL) as a fractional reference model. The desired response of the fractional order plant can be achieved by choosing an appropriate reference model with suitable parameters μ and ω_u.

The simulation results show that the proposed concept FRIT-FPID is a suitable way for the control of the fractional order systems.

With the proposed methodology we get closed loop fractional order system robust to gain variations and measurement noise.

The FRIT-FPID method is especially oriented to practical implementation thanks to online parameters tuning of fractional order controller using RLS and the DAFI filter as a powerful tool for numerical simulation of fractional operators.

Future research direction is to investigate the effectiveness of the proposed technique by carrying out its practical implementation.

REFERENCES

[1] Souma, S., Kaneko, O., and Fujii, T. (2004). A new method of controller parameter tuning based on input output data, –fictitious reference iterative tuning. *Proceedings of IFAC Workshop on Adaptation and Learning in Control and Signal Processing*, 789–794.

[2] Tasaka, K., Kano, M., Ogawa, M., Masuda, S., and Yamamoto, T. (2009). Direct PID tuning from closed loop data and its application to unstable processes. *Transaction of the Institute of Systems, Control and Information Engineers*, 22(4), 137–144 (in Japanese).

[3] Masuda, S. (2010). Adaptive PID control based on the online FRIT approach.*Proceedings of SICE 10th Annual Conference on Control Systems*, 16532 (in Japanese).

[4] Wakasa, Y., Kanagawa, S., Tanaka, K., and Nishimura, Y. (2011). FRIT for systems with dead-zone and its application to ultrasonic motors. *IEEJ Transactions on Electronics, Information and Systems*, 131(6), 1209–1216

[5] Yamashina, Y., Kaneko, O., and Yamamoto S. (2011). Real-time parameter tuning for the feed-forward controller in a two-degree-of-freedom control system with FRIT.*Proceedings of the 10th SICE Control Division Conference*, 18412 (in Japanese).

[6] Miller, K.S., Ross, B.: An Introduction to the Fractional Calculus and Fractional Differential Equations.Wiley, New York (1993)

[7] Sabatier, J., et al. (ed.): Advances in Fractional Calculus: Theoretical Development and Applications in Physics and Engineering. Springer, Dordrecht (2007)

[8] E. Reyes-Melo, J. Martinez-Vega, C. Guerrero-Salazar, and U. Ortiz-Mendz, "Application of fractional calculus to the modeling of dielectric relaxation phenomena in polymeric materials," J. Appl. Polym. Sci., Vol. 98, No. 2, pp. 923-935, Jul. 2005.

[9] N. Bertrand, J. Sabatier, O. Briat, and J. Vinassa, Fractional non-linear modelling of ultracapacitors," Commun. Nonlinear Sci. Numer. Simulat., Vol. 15, No. 5, pp. 1327-1337, May 2010

[10] K. B. Oldham and J. Spanier, The fractional calculus: theory and application of differentiation and integration to arbitrary order, Springer, 2006.

[11] Sabatier, O. P.Agrawal and J. A. Tenreiro Machado, Advances in Fractional Calculus: Theoretical Development and Applications in Physics and Engineering. Springer, (2006)

[12] Podlubny, I.: Fractional order systems and $PI^{\lambda}D^{\mu}$controllers. IEEE Trans. Autom. Control 44(1), 208–214 (1999)

[13] F. C. Meral, T. J. Royston, and R. Magin, "Fractional calculus in viscoelasticity: an experimental study," Commun. Nonlinear Sci. Numer. Simulat., Vol. 15, No. 4, pp. 939-945, Apr. 2010.

[14] Haykin, S. (2002). *Adaptive Filter Theory*.Prentice-Hall, Upper Saddle River, NJ, 4th edition.

[15] A. Djouambi, A. Charef and A. Voda.(2013). Numerical Simulation and Identification of Fractional Systems using Digital Adjustable Fractional Order Integrator.European Control Conference, Zurich, Switzerland. 2013

[16] Yuji Wakasa,Kanya Tanaka, Yuki Nishimura. Online Controller Tuning via FRIT and Recursive Least-Squares.IFAC Conference on Advances in PID Control. Brescia ,2012.

[17] A. Oustaloup, C. Olivier, L. Ludovic, *Représentation et identification par modèle non entier, LAVOISIER,* 2005.

[18] M. A. Al-Alaoui, M.A, " Novel IIR differentiator from the Simpson Integration rule," *IEEE Transactions on Circuits and Systems I. Fundamental Theory and Applications,* 41 (2), 1994, pp 186-187.

[19] B. M. Vinagre, I. Podlubny, A. Hernandez, V. Feliu, "Some approximations of fractional order operators used in control theory and applications," *Fractional Calculus & Applied Analysis*, 3(3), 2000, pp. 231–248.

[20] I. Petras, I. Podlubny, P. O'Leary, L. Dorcak, "Analogue fractional order controllers: Realization, tuning and implementation, *Proceedings of the ICCC'2001*, 2001, pp. 9 - 14, Krynica,Poland.

[21] Y. Q. Chen, and Kevin L. Moore, "Discretization Schemes for Fractional-Order Differentiators and Integrators.*IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications*, 49(3): 2002, pp 363-367.

[22] A. Charef and D. Idiou, "Design of analog variable fractional order differentiator and integrator, *Nonlinear Dyn,* DOI 10.1007/s11071- 012-0370-x, Springer Science+Business Media B.V. 2012.

[23] Charef A., H.H. Sun , *Fractional System as represented by singularity function,* IEEE Trans. on Auto. Contr., vol. 37, n° 9, 1992.

[24] Miller K.S., Ross B., *An introduction to the fractional calculus and fractional differential equations,* John Wiley and Sons, 1993.

[25] Podlubny I., Fractional differential equations: an introduction to fractional derivatives, fractional differential equations, to methods of their solution and some of their applications, San Diego: Academic Press, 1999.

[26] Shiro Masuda, Manabu Kano, Yusuke Yasuda and Guo-Dong Li. (2010). A fictitious reference iterative tuning method with Simultaneous delay parameter tuning of the reference model. International Journal of Innovative Computing, Information and Control.

[27] Oustaloup A., (1991). La commande CRONE, Paris: Hermès.

[28] Barbosa, R. S., J.A.T. Machado, and I.M. Ferreira, (2004). Tuning of PID controller based on Bode's ideal transfer function, Nonlinear Dynamics, vol. 38, pp. 305-321.

[29] A. Djouambi, A. Voda, A. Charef, (2012), Estimation Récursive des Paramètres d'un Contrôleur |PIαDβ Robuste d'Ordre Non Entier, Conférence Internationale Francophone d'Automatique (CIFA2012), Grenoble, France. 4-6 Juillet 2012

[30] Oustaloup A, F. Leveron, B. Mathieu and F. M. Nanot, Frequency-band complex non integer differentiator: Characterization and synthesis, IEEE, Trans. On circuit and systems I: Fundamental theory and applications, Vol. 47, no. 1, 25-39, 2000.

[31] Astrom, K. and Hagglund, T., *PID Controllers: Theory, Design, and Tuning*, Instrument Society of America, North Carolina, 1995.

[32] Celaleddin Yeroglu, usret Tan, (2010), Optimal PID Controller Design Based on Fractional Order Reference Model, IFAC 2010.

[33] Ramiro S. Barbosa, Manuel F. Silva, and J. A. Tenreiro Machado, (2009), Tuning and Application of Integer and Fractional Order PID Controllers,

ISSN 1737-9334

Intelligent Engineering Systems and Computational Cybernetics, Springer Science+Business Media B.V. 2009.

[34] Ivo Petras, (2009), FRACTIONAL ORDER FEEDBACK CONTROL OF A DC MOTOR, Journal of ELECTRICAL ENGINEERING, VOL. 60, NO. 3, 2009, 117–128.

[35] A. Djouambi, A. Charef, and A. Voda, (2010), FRACTIONAL ORDER CONTROLLER BASED ON BODE'S IDEAL TRANSFER FUNCTION, Control and Intelligent Systems, Vol. 38, No. 2, 2010.

[36] Bode H. W., (1945). Network Analysis and Feedback Amplifier Design, Van Nostrand, New York.

[37] A. Le Méhauté and G. Crepy, "Introduction to transfer and motion in fractal media: The geometry of kynetics," *Solid State Ionics*, no. 9–10, pp. 17–30, 1983.

[38] S. Westerlund, Dead matter has memory! , Kalmar, Sweden: Causal Consulting, 2002.

[39] F. Boulgameh, A. Djouambi and M. Remram, (2014), Small Signal Fractional Order Modeling of PN Junction Diode. The Second International Conference on Electrical Engineering and Control Applications ICEECA'2014, Constantine, Algeria. 2014.

[40] YangQuan Chen, Ivo Petras and Dingyu Xue, Fractional Order Control - A Tutorial, 2009 American Control Conference Hyatt Regency Riverfront, St. Louis, MO, USA, June 10-12, 2009.

[41] OLDHAM, K. B.—SPANIER, J. The Fractional Calculus, Academic Press, NY; 1974.

[42] Abdelbaki Djouambi. 2008, CONTRIBUTION A LA COMMANDE CRONE. Phd thesis, Université Mentouri de Constantine Algérie.

[43] Igor Podlubny, Fractional-Order Systems and $PI^{\gamma}D^{\mu}$ Controllers, IEEE TRANSACTIONS ON AUTOMATIC CONTROL, VOL. 44, NO. 1, JANUARY 1999

ISSN 1737-9334

Investigation of Tuning Parameters Selection in Predictive Functional Control

Neslihan Karaş[#1], Leyla Gören Sümer[*2]

[#]*Department of Control and Automation Engineering, Yildiz Technical University*
Istanbul, Turkey
[1]nkaras@yildiz.edu.tr

[*]*Department of Control and Automation Engineering, Istanbul Technical University*
Istanbul, Turkey
[2]gorenl@itu.edu.tr

***Abstract*— This paper gives an investigation of tuning parameters selection in predictive functional control (PFC). PFC is a kind of model predictive control (MPC). With the algorithm based on a single coincidence horizon, it has low calculation time compared to other model predictive control types. In addition, it differs from classical control methods in terms of physical parameters. It is a very advantageous method due to its ability to handle large dead times, constraints. It has implicit integral effect so it has not disadvantage of integral wind-up as in PID and the steady state error is zero. So, predictive functional control has a very significant penetration into markets where classical methods predominate.**

PFC is a method that has advantages as well as disadvantages. One of these disadvantages is that the choice of coincidence horizon is determined by trial and error. Therefore, the choice of the coincidence horizon is examined for first order systems and higher order systems in this study. Another tuning parameter of the PFC is the desired pole. This choice is intuitive. A desired pole is chosen according to the target trajectory.

***Keywords*— Prediction, model predictive control, predictive functional control, tuning**

I. Introduction

The principles of PFC were established by J. Richalet in 1968, and the first applications were made in 1970s. Today, PFC is a widely used method in the industry [1].

MPC is widely used in the industry because it can work without expert intervention for a long time. PFC is an MPC type characterized by a very simple calculation algorithm and excellent control performance. It has a simpler algorithm than other MPC types. Thus, PFC is cheap to implement. The simplicity of the PFC with a constant reference signal comes from the assumption that the control signal is constant during the control horizon and that the output is optimized only at a single point in the future. Thus, PFC does not require numerical minimization or matrix inversion.

After the introduction, the rest of the paper is organized as follows. Section 2 introduces model-based prediction and how to formulate prediction equations. Section 3 is the PFC section. In this section, basic concepts and mathematical definition of PFC are mentioned. Control law formulation is found for first order and higher order systems. Section 4 describes how the tuning parameters are chosen for first order and higher order systems. As a result, the paper is concluded in section 5.

II. Model Based Predictive Control

A classical controller observes only the current process variables and remembers the past variables, while the predictive controller also observes future process variables. Therefore, it is very logical to use the predictions of the output in determining a control strategy [2].

In the model-based prediction, the predictions of the output are obtained through the model [1]. The independent model is used in PFC as shown in Fig. 1.

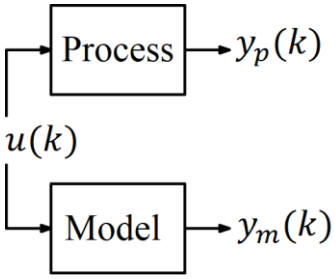


Fig. 1 Independent model

The following steps are followed to write the prediction equations of the output of a system [3], [4].

Step 1: Firstly, a system is considered as like (1).

$$G(z) = \frac{y(k)}{u(k)} = \frac{b(z)}{a(z)} \tag{1}$$

Step 2: Equation (1) becomes as like (2) with cross-multiplication:

$$a(z)y(k) = b(z)u(k) \tag{2}$$

Step 3: Unbiased prediction can be ensured by using (3).

$$\begin{aligned} &\underbrace{a(z)\Delta(z)}_{A(z)}\, y(k) = b(z)\,\underbrace{\Delta(z)u(k)}_{\Delta u(k)} \\ &A(z) = 1 + A_1 z^{-1} + \ldots + A_{n+1} z^{-n-1} \\ &b(z) = b_0 + b_1 z^{-1} + \ldots + b_n z^{-n} \\ &\Delta u(k) = u(k) - u(k-1) \end{aligned} \tag{3}$$

Step 4: Prediction equations are created from 1 to coincidence horizon n_y.

ISSN 1737-9334

$$
\begin{aligned}
&\underbrace{a(z)\Delta(z)}_{A(z)}\, y(k) = b(z)\,\underbrace{\Delta(z)u(k)}_{\Delta u(k)} \\
&y(k+1) + A_1 y(k) + \cdots + A_{n+1} y(k-n) \\
&\qquad = b_1 \Delta u(k) + \cdots + b_{n-1}\Delta u(k-n+1) \\
&y(k+2) + A_1 y(k+1) + \cdots + A_{n+1} y(k-n+1) \\
&\qquad = b_1 \Delta u(k+1) + \cdots + b_n \Delta u(k-n+2) \\
&\qquad \vdots \\
&y(k+n_y) + A_1 y(k+n_y-1) + \cdots + A_{n+1} y(k+n_y+1-n) \\
&\qquad = b_1 \Delta u(k+n_y-1) + \cdots + b_n \Delta u(k+n_y-n)
\end{aligned}
\tag{4}
$$

Step 5: If the prediction equations are written in vector matrix form, then equation (5) is obtained.

$$
\begin{aligned}
&\underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ A_1 & 1 & \cdots & 0 \\ A_2 & A_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}}_{C_A}
\underbrace{\begin{bmatrix} y(k+1) \\ y(k+2) \\ \vdots \\ y(k+n_y) \end{bmatrix}}_{\vec{y}_{k+1}}
+ \underbrace{\begin{bmatrix} A_1 & A_2 & \cdots & A_{n+1} \\ A_2 & A_3 & \cdots & 0 \\ A_3 & A_4 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}}_{H_A}
\underbrace{\begin{bmatrix} y(k) \\ y(k-1) \\ \vdots \\ y(k-n) \end{bmatrix}}_{\overleftarrow{y}_k} \\
&\qquad = \underbrace{\begin{bmatrix} b_1 & 0 & \cdots & 0 \\ b_2 & b_1 & \cdots & 0 \\ b_3 & b_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}}_{C_{zb}}
\underbrace{\begin{bmatrix} \Delta u(k) \\ \Delta u(k+1) \\ \vdots \\ \Delta u(k+n_y-1) \end{bmatrix}}_{\Delta \vec{u}(k)} \\
&\qquad + \underbrace{\begin{bmatrix} b_2 & b_3 & \cdots & b_n \\ b_3 & b_4 & \cdots & 0 \\ b_4 & b_5 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}}_{H_{zb}}
\underbrace{\begin{bmatrix} \Delta u(k-1) \\ \Delta u(k-2) \\ \vdots \\ \Delta u(k-n) \end{bmatrix}}_{\Delta \overleftarrow{u}(k-1)}
\end{aligned}
\tag{5}
$$

Step 6: So, output prediction $\vec{y}_{k+1}$ is expressed as in (6).

$$
\begin{aligned}
&\vec{y}_{k+1} = H\Delta\vec{u}_k + P\Delta\overleftarrow{u}_{k-1} + Q\overleftarrow{y}_k \\
&H = C_A^{-1} C_{zb} \\
&P = C_A^{-1} H_{zb} \\
&Q = -C_A^{-1} H_A
\end{aligned}
\tag{6}
$$

As an important note, future inputs are considered constant in PFC as shown in (7).

$$u(k) = u(k+1) = u(k+2) = \cdots \tag{7}$$

PFC has low computational time than other MPC types because PFC uses only n_y.line of these matrices.

III. Predictive Functional Control

In this section, firstly, the basic concepts of PFC and its mathematical definition are mentioned. From this mathematical definition, the control law for a first order system is found. Then the PFC control law for higher order systems is found.

A. Basic Concepts of PFC

There are 3 basic concepts in PFC. These are [5]:

- Target trajectory
- Coincidence horizon
- Basis functions

1) Target Trajectory: Richalet recommended a reference trajectory that approaches the reference with a first order exponential function [2]. The reference trajectory r^{tr} in the domain s is given in (8) and in the domain z in (9) where r is set-point.

$$r^{tr}(s) = \frac{e^{-sT_d}}{\tau_r s + 1} r \tag{8}$$

$$r^{tr}(z) = z^{-d_m} \frac{(1-\lambda)z^{-1}}{1-\lambda z^{-1}} r \tag{9}$$

The target trajectory is determined by the discrete pole λ. As shown in (10) and (11), the discrete pole can be determined by reference time constant τ_r or settling time $CLTR$ where T_s is sampling time.

$$\lambda = e^{-\frac{T_S}{\tau_r}} \tag{10}$$

$$\lambda = e^{-\frac{3T_S}{CLTR}} \tag{11}$$

2) *Coincidence Horizon:* The control aim is to reach the reference trajectory after n_y step. The minimum value of the coincidence horizon is 1. Coincidence horizon has positive integer values [3].

The target trajectory's discrete pole (λ) and coincidence horizon (n_y) are the tuning parameters of the PFC.

3) Basis Functions: The vector of future $u(k)$ is not established directly. Instead, the projection μ_j of $u(k)$ is determined onto a finite set of basic functions [1].

$$u(k+j) = \sum \mu_j F_j(i) \;\; j = 0,,\ldots,N-1 \;\; 0 \le i \le n_y \tag{12}$$

In PFC, the reference can be step, ramp or parabola. The shape of the reference determines the number of coincidence horizons. For the step reference, there is one unknown variable as in (13). Therefore, there is only one coincidence horizon.

$$u(k) = \mu_0 \tag{13}$$

For ramp reference there is two unknown variable as in (14). Therefore, two prediction equations and two coincidence horizon are used.

$$u(k) = \mu_0 + \mu_1 i \tag{14}$$

For parabola reference there is three unknown variable as in (15). Therefore, three prediction equations and three coincidence horizon are used.

$$u(k) = \mu_0 + \mu_1 i + \mu_2 i^2 \tag{15}$$

In this study, there is only one coincidence horizon because the reference is considered as step.

B. Mathematical Definition of PFC

The basic philosophy is based on the coincidence of the reference trajectory and the process output at the n_y. At this point, the desired change in the reference trajectory Δr^{tr} is equal to the desired change in the process output Δy_p. Furthermore, PFC assumes that the predicted change of the model Δy_m is equal to the change of process.

As a result, the desired change in the process output is calculated from the desired change in the reference trajectory

ISSN 1737-9334

and the predicted change of the model output [2]. Equations (16) and (17) are obtained from these assumptions.

$$\Delta r^{tr} = \Delta y_p \tag{16}$$

$$\Delta y_p = \Delta y_m \tag{17}$$

Fig. 2 shows the PFC strategy. From Fig. 2, equations (18)-(22) are written.

$$y_p(k+n_y) = r^{tr}(k+n_y) \tag{18}$$

$$\Delta y_p = y_p(k+n_y) - y_p(k) \tag{19}$$

$$\Delta y_m = y_m(k+n_y) - y_m(k) \tag{20}$$

$$\varepsilon(k) = r - r^{tr}(k) = r - y_p(k) \tag{21}$$

$$\Delta r^{tr} = \varepsilon(k) - \varepsilon(k+n_y) \tag{22}$$

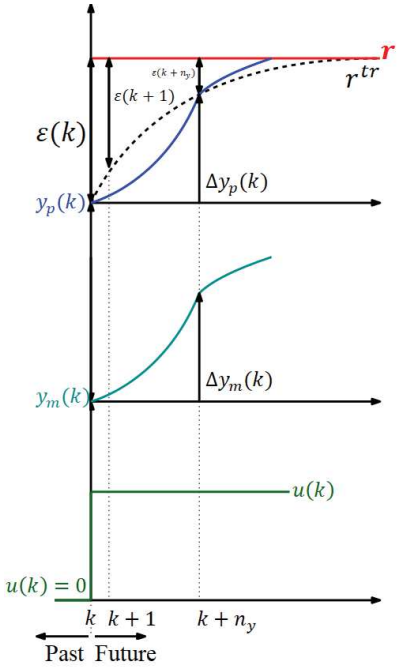


Fig. 2 PFC strategy

It is assumed that the difference between the set-point and the reference trajectory decreases by λ in each sample when changing the reference trajectory. This assumption is shown in (23).

$$\varepsilon(k+n_y) = \lambda^{n_y}\varepsilon(k) \tag{23}$$

By substituting (21) and (23) in (22), the reference trajectory change is obtained as in (24).

$$\Delta r^{tr} = (1-\lambda^{n_y})(r - y_p(k)) \tag{24}$$

Mathematical definition is found by equating (20) and (24).

$$(1-\lambda^{n_y})(r - y_p(k)) = y_m(k+n_y) - y_m(k) \tag{25}$$

C. PFC for first order systems

From section II, n_y step ahead prediction equation for a first order system (26) can be easily, directly written (27).

$$G(z) = \frac{bz^{-1}}{1+az^{-1}} \tag{26}$$

$$y(k+n_y) = (-a)^{n_y}y(k) + b\frac{1-(-a)^{n_y}}{1+a}u(k) \tag{27}$$

If the model's prediction equation (27) is written instead of (25), formulation of control law is obtained as in (28) [5].

$$u(k) = \frac{(1-\lambda^{n_y})\left(r-y_p(k)\right)+(1-(-a_m)^{n_y})y_m(k)}{\frac{b_m}{1+a_m}(1-(-a_m)^{n_y})} \tag{28}$$

If the coefficients in (28) are called k_0 and k_1, the formulation of control law is written as in (29).

$$u(k) = k_0\left(r - y_p(k)\right) + k_1 y_m(k) \tag{29}$$

D. PFC for higher order systems

It is possible to find a control law for higher order systems by using prediction equation in section II. In the mathematical definition of PFC, the prediction equation is written instead of $y_m(k+n_y)$ [5].

Since PFC selects a single coincidence horizon for step references, only the coincidence horizon line is considered from the vector matrix format in (6). In this case, the prediction equation of $y_m(k+n_y)$ is written as in (30) . H_{n_y}, P_{n_y}, Q_{n_y} is n_y. line of H, P, Q matrices.

$$y_m(k+n_y) = H_{n_y}\underset{\rightarrow}{\Delta u}(k) + P_{n_y}\underset{\leftarrow}{\Delta u}(k) + Q_{n_y}\underset{\leftarrow}{y_m}(k) \tag{30}$$

Because of the assumption that future inputs is constant (7), the sum of the elements of H_{n_y} is needed as in (31).

$$h_{n_y} = \sum(H_{n_y}) \tag{31}$$

As a result, the prediction equation in (30) is substituting in mathematical definition, and the control law is found as in (32).

$$u(k) = u(k-1) + \frac{(1-\lambda^{n_y})\left(r-y_p(k)\right)+P_{n_y}\underset{\leftarrow}{\Delta u}(k-1)-Q_{n_y}\underset{\leftarrow}{y_m}(k)+y_m(k)}{h_{n_y}} \tag{32}$$

IV. Choice of Tuning Parameters

A. Choice of Tuning Parameters for First Order Systems

Choice of λ is intuitive and coincidence horizon is determined by trial and error. However, there is a fact in the literature that; The best desired answer for first order systems is provided by n_y = 1 [3]. To prove this, the closed loop polynomial is used. The actual closed loop pole of the system for the nominal state $y_p = y_m$ is obtained by following the steps below.

Step 1: Consider the first order discrete system as in (26).

Step 2: The model output is written instead of in the (28). So the closed-loop polynomial is found as in (33) where δ is a scalar.

$$\left(1 + \left(a + \frac{((-a)^{n_y}-\lambda^{n_y})\,(1-(-a)^{n_y})b}{b\left(\,(1-(-a)^{n_y})\right)}\right)z^{-1}\right)y(k) = \delta r \tag{33}$$

Step 3: The pole is found.

$$Pole = -a - \frac{((-a)^{n_y}-\lambda^{n_y})\,(1-(-a)^{n_y})}{(1-(-a)^{n_y})} \tag{34}$$

ISSN 1737-9334

The pole where the coincidence horizon is one is obtained by writing $n_y = 1$ in (34). In this case, the actual closed loop pole is equal to the desired pole. The pole where the coincidence horizon goes to infinity is obtained by writing $n_y \rightarrow \infty$ in (35). In this case, the actual closed loop pole is equal to the open loop pole.

As evidenced by (34), in a first-order model with coincidence horizon $n_y = 1$, the desired closed-loop pole is exactly achieved in nominal state. For a first-order system with a large coincidence horizon, the actual closed-loop pole leads to the open-loop pole, and therefore the tuning parameter λ is meaningless.

For a system with an open pole of 0.8 as in (35), for the various λ actual closed-loop poles depending on the coincidence horizon are shown in Fig. 3.

$$G_1(z) = \frac{0.25z^{-1}}{1-0.8z^{-1}} \tag{35}$$

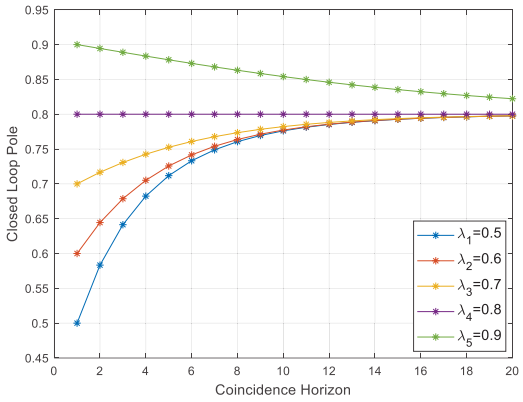


Fig. 3 Actual closed loop pole dependence on n_y and λ for G_1

As shown in Fig. 3, the actual closed-loop pole is exactly equal to the desired pole when the coincidence horizon is one. As the coincidence horizon grows, the curve moves to the open loop pole.

To clearly observe the effect of the coincidence horizon, the closed loop responses for the various coincidence horizon of $G_1(z)$ is plotted as in Fig. 4. The desired pole is 0.7. As n_y grows, the response go away from the target trajectory.

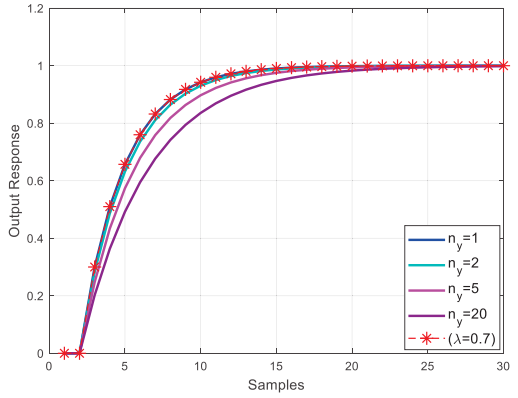


Fig. 4 Closed loop responses with for various n_y for G_1

By choice the target trajectory, λ is actually chosen. That is, the choice of λ is chosen according to which closed loop time constant or settling time is desired. Therefore, the tuning of λ is intuitive and optional. This choice is not dependent on the coincidence horizon.

The target trajectories depending on the various λ are shown in Fig. 5. Fig. 5 shows the trajectories of the first order pole at 0.5, 0.7 and 0.9 poles. As can be seen, the process output is on trajectory of pole 0.7 as desired for $n_y = 1$.

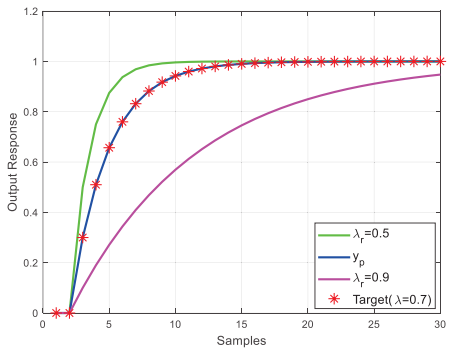


Fig. 5 Dependence on choice of various λ with $n_y = 1$ for G_1

B. Choice of Tuning Parameters for Higher Order Systems

In order to examine the coincidence horizon choice, a higher order system is considered (36).

$$G_2(z) = \frac{0.02z^{-1}+0.04z^{-2}}{1-1.4z^{-1}+0.45z^{-2}} \tag{36}$$

Contrary to the one proposed for first order systems, choice of the coincidence horizon as one is not recommended for higher order. Because in systems with unstable zeros, the coincidence horizon cannot be selected as one [6]. This case is shown Fig. 6 and Fig. 7.

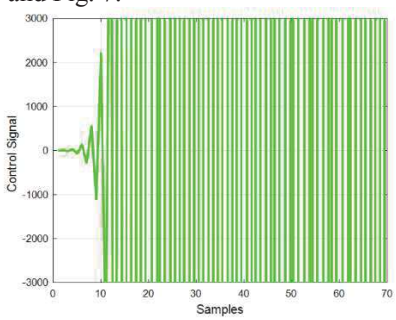


Fig. 6 G_2's control signal for $n_y = 1$

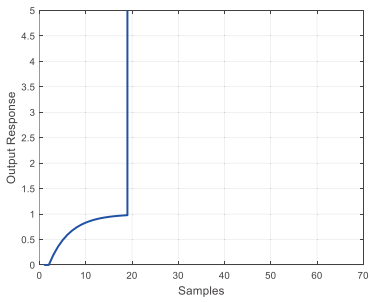


Fig. 7 G_2's output signal for $n_y = 1$

ISSN 1737-9334

Actual closed loop poles for G_2 with various choice of n_y are shown in Fig.8.

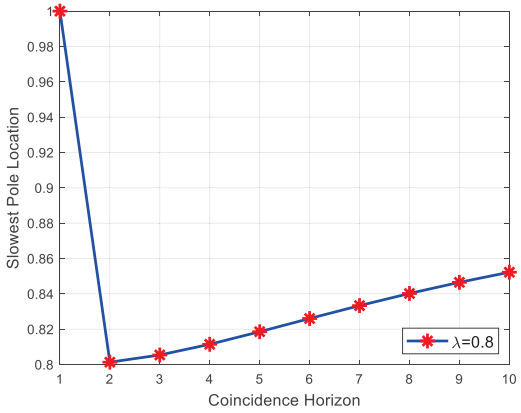


Fig. 8 Actual closed loop pole for G_2 with various choice of n_y

The coincidence horizon should be $2 \leq n_y \leq 4$ according to Fig.8 in order to be close to the desired pole.

There is a suggestion in [3], [6] that; For systems with monotone convergent behavior after immediate transitions, a good coincidence horizon may be selected as the range in where the open loop response is risen to around 40% to 80% of the steady state and the gradient is significant.

If this suggestion is used for the G_2 system, Fig. 9 is obtained.

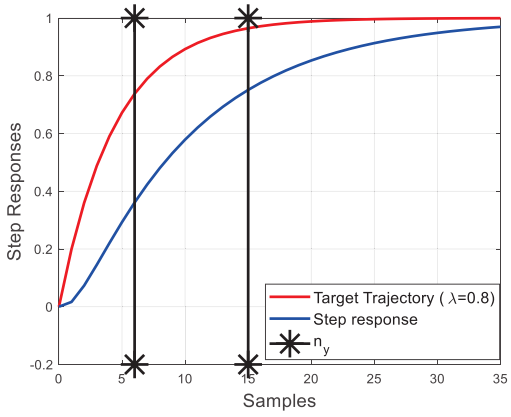


Fig. 9 n_y range for G_2

As shown in Fig. 9, the coincidence horizon should be $6 \leq n_y \leq 15$ from the 40% -80% rule.

For $2 \leq n_y \leq 4$, control signals are shown in Fig. 10 and the output signals are shown in Fig. 11. For $6 \leq n_y \leq 15$, control signals are shown in Fig. 12 and the output signals are shown in Fig. 13.

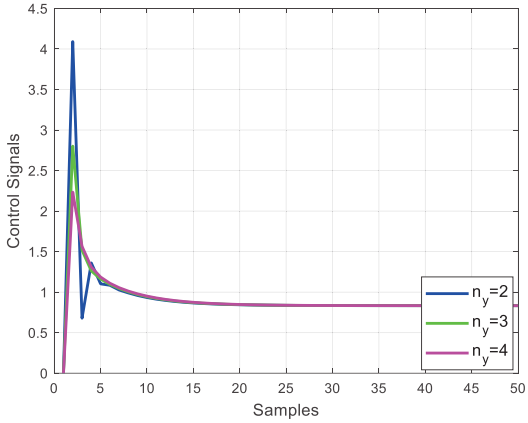


Fig. 10 G_2's control signals for $n_y = 2,3,4$

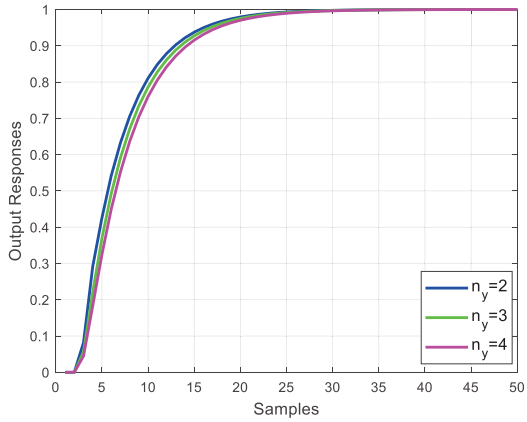


Fig. 11 G_2's output signals for $n_y = 6,10,15$

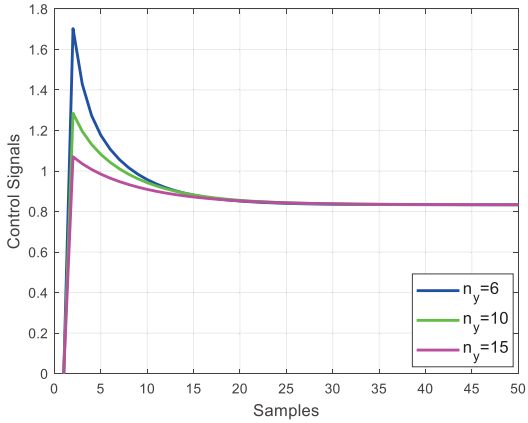


Fig. 12 G_2's control signals for $n_y = 2,3,4$

ISSN 1737-9334

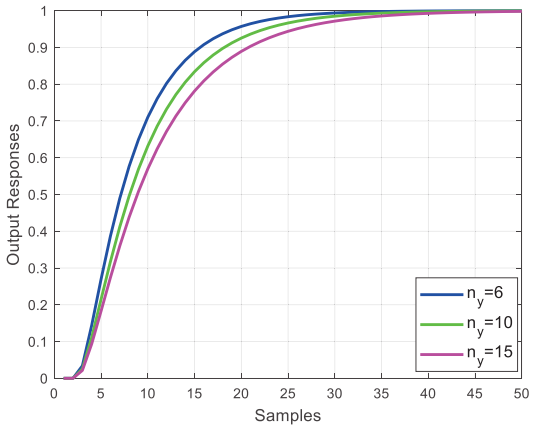


Fig. 13 G_2's output signals for $n_y = 6,10,15$

As seen in Fig.10 and Fig 11, output responses are almost in the desired pole trajectory. As the n_y becomes smaller, a more active control signal is obtained.

As seen in Fig.12 and Fig 13, within the recommended range by 40% -80% the closed-loop behavior is good in all cases, although as expected, the desired speed-up of G_2 is not achieved due to the need to use high n_y.

V. Conclusions

The tuning of PFC is not obvious and systematic. The selection of tuning parameters is determined by trial and error. However, in this case an infinite choices are available. Therefore the effect of the tuning parameters should be examined. In this study, the selection of the parameters for a first order system and a higher order system is investigated. For this purpose, firstly the closed loop poles depending on the coincidence horizon is plotting for first order systems. As shown in this drawing, PFC obtains the closest response to the desired pole for first order systems when $n_y = 1$. When $n_y \rightarrow \infty$, the target trajectory converges open loop behavior.

For the higher order systems, the selection is more complicated. Intuitively, the core failing is the desire to follow a first order trajectory, and using a constant future input as this is clearly unrealistic for systems with complex dynamics. Consequently, some suggestions were made as to how one could determine an appropriate range of n_y and λ before any fine tuning.

First observation is that with unstable poles, $n_y = 1$ cannot be applied. Therefore, it is reasonable to determine a n_y range. For the higher order systems open loop step response is examined to determine a logical n_y range. With this suggestion the desired speed-up of G_2 is not achieved but the closed-loop behavior is good in all cases.

To sum up, small coincidence horizon ensures that there is a good link between the λ and the dominant closed-loop pole. However, small coincidence horizon can leads to poor transient behavior. As n_y decreases, it causes an over-active control signal. As n_y grows, the distancing from the target trajectory occurs.

For higher order systems, the tuning rules and efficacy of the parameters λ and n_y are not consistent and vary from one example to another depending on the underlying process dynamics. Extending this study for more challenging dynamics is a future work.

References

[1] J. Richalet, D. O'Donavan, *Predictive Functional Control: Principles and Industrial Applications*, Springer-Verlag, 2009.

[2] R. Haber, R. Bars, U. Schmitz, *Predictive Control in Process Engineering: From The Basics To The Applications*, Wiley-VCH, 2011.

[3] J. A. Rossiter, R. Haber, "The effect of coincidence horizon on predictive functional control," Processes, 3, pp. 25–45, 2015.

[4] J. A. Rossiter "Notes on Multi-Step Ahead Prediction Based on the Principle of Concatenation. Proceedings of the Institution of Mechanical Engineers" Part I: Journal of Systems and Control Engineering, 207(4), pp. 261–263, 1993.

[5] R. Haber, J.A. Rossiter & K. Zabet "An alternative for pid control:Predictive functional control-a tutorial," *American Control Conference (ACC), IEEE,* pp. 6935-6940, 2016.

[6] J.A. Rossiter, *A First Course in Predictive Control*, 2nd ed. London, UK,CRC Press, 2018.

ISSN 1737-9334

Perturb & Observe based Integer Order Sliding Mode MPPT Control of Solar Photovoltaic

[1,2]Hina Gohar Ali
[1]Department of Telecommunications and Systems Engineering
Autonomous University of Barcelona
08193 Bellaterra, Barcelona
hina.goharali@e-campus.uab.cat
[2]School of Electrical Engineering and Computer Science
National Univ. of Science and Technology
44000, Islamabad, Pakistan
hinagohar1302@gmail.com

[3] Jorge Herrera
[3]Departamento de Ingeniera
Facultad de Ciencias Naturales e Ingeniera
Universidad de Bogot Jorge Tadeo Lozano
Bogot, Distrito Capital, Colombia
jorgea.herrerac@utadeo.edu.co

[1]Ramon Vilanova
[1]Department of Telecommunications and Systems Engineering
Autonomous University of Barcelona
08193 Bellaterra, Barcelona
Ramon.Vilanova@uab.cat

[4]Syed Attique Shah
[4]Informatics institute, Istanbul Technical University
34469 Istanbul, Turkey
Shah@itu.edu.tr

***Abstract*—This paper presents a method for tracking the maximum power point (MPPT) of an array of photovoltaic panels under varying radiation and temperature by applying an integer order sliding mode controller with perturb & observe (P&O) MPPT algorithm. P&O algorithm is simulated in PSIM to generate the reference and sliding mode controller is designed in Matlab Simulink to track the reference provided by P&O MPPT. SMC controller is than coupled through simcoupler interface between PSIM and Matlab Simulink. The results from simulation highlight that this method can track the maximum power point quickly and accurately under standard conditions and sudden changes in illumination intensity, and has faster tracking speed and stable characteristics inherent in fixed duty cycle.**

***Index Terms*—maximum power point tracking (MPPT), photovoltaic, sliding mode controller, perturb & observe**

I. Introduction

With the acceleration of economic globalization and the rapid development of industrial economy, the worlds energy shortage and environmental pollution have become two important factors that restrict the sustainable development of human society [1]. Photovoltaic (PV) power generation is a very important form of energy conversion, and efficiency is one of the key issues that constrain its development. To operate the photovoltaic system at maximum power point (MPP) the power electronics components of the photovoltaic cell needs to be adjusting. Temperature and solar irradiance are the main factors affecting the PV array maximum power. Maximum power point tracking (MPPT) method is one of the most well-known and widely used for the extraction of the maximum power of the PV system which has always been applied in photovoltaic energy conversion system [2]. In literature, the most commonly used MPPT methods are : Perturb and Observe (P&O) [3] and Incremental Conductance (IC) [4] which are the most widely used and popular MPPT algorithms. Other MPPT proposed methods are the Constant Voltage Tracking (CVT), Fractional Open-circuit Voltage (FOVC), Fractional Short Circuit Current (FSC), Fuzzy control methods [5] and so on. There are numerous controllers available which can also be used along with MPPT algorithms. The purpose of these controllers is to guarantee a proper tracking of the reference voltage generated by the MPPT algorithm for maximum possible power extraction [6]. PV module has non-linear characteristics and switching converters used in PV system has non-linear behaviour also. Non-linear controllers are considered widely in PV system to model non-linear dynamics which has gained considerable attention. Among non-linear controllers Sliding mode control (SMC) has attracted a lot of interest due to its fast response speed, robustness and non-linearity [7]. Sliding mode control is a nonlinear control strategy based on the variable structure theory. The contribution of this paper is designing of an integer order SMC for the extraction of maximum power from PV panel under variable irridance and temperature. P&O MPPT algorithm is employed to generate the reference PV voltage for SMC controller which is simulated in PSIM. The SMC controller is derived based on nonlinear mathematical model of PV Panel with boost converter. The robust stability of the proposed controller is assured by Lyapunov theory. Implementation of SMC controller is performed in Matlab Simulink which is provided as control signal to PV MPPT in PSIM through Simcoupler interface between Simulink and Matlab.

This work has received financial support Spanish Government under Project DPI2016-77271-R co-funded with European Union ERDF funds.

ISSN 1737-9334

II. Characteristics of Photovoltaic System

The proposed PV system circuit schematic is shown in Fig.1, which is composed of a PV module and a boost converter. Photovoltaic panel is used as power source to supply power to this standalone PV system. The photovoltaic panel model based on two diodes is used which involves identification of more parameters and better efficiency. Two-diode model gives more accurate I-V and P-V output curve characteristics as compared to single diode model [8]. In

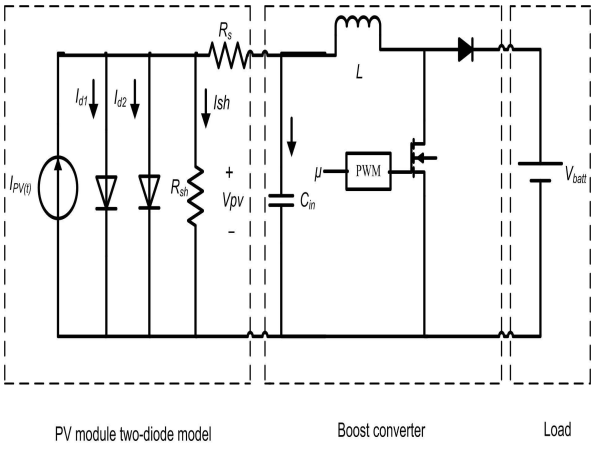


Fig. 1. Circuit diagram of the proposed PV system.

this section, a photovoltaic module is introduced which consists of photovolatic cells connected in parallel and series combination. Photovoltaic cells are devices that convert solar energy into electrical energy directly through photoelectric effect. To represent the current-voltage (I-V) characteristics photovoltaic cells single-iode model and two-diode model are a common representation available in the literature of the solar cell behaviour. The equivalent circuit of phovoltaic panel based on two-diode model consists of a seris resistor (R_s), shunt resistor (R_{sh}), two diodes in anti-parallel and a current source. Two-diode model takes into the consideration of recombination effect by including second diode in parallel with the single cell model. It is basically the modified form of single diode model. The following characteristic Eq. 1 shows the relationship between the solar panel current and voltage.

$$I = I_{ph} - I_{o1}\left(e^{\frac{V+IR_s}{N_{s1}V_T}} - 1\right) - I_{o2}\left(e^{\frac{V+IR_s}{N_{s2}V_T}} - 1\right) - \frac{V+IR_s}{R_p} \quad (1)$$

In Eq. 1, V & I is the voltage and output current of solar cell, I_{ph} is the light generated current, I_s is the reverse saturation current of the diode, N_s is the number of cells in series, Io is the diode saturation current, V_T is thermal voltage equivalent while R_s and R_{sh} is the series and shunt resistances respectively.

III. MPPT Algorithm

To extract maximum power from photovoltaic system MPPT technique is used in power electronic circuits. Dc-Dc voltage converter is generally used as an interface between the PV panel and load to extract the maximum power of the panel. Dc-Dc boost converter is chosen in this PV system design configuration. To tune the input impedance of the boost converter duty cycle is generated by pulse width modulation (PWM). The boost converter has the tendency to change the operating point of PV module and can maximize the output power. Perturbation and Observe (P&O) MPPT algorithm is the most commonly used method for controlling Dc-Dc converter due to its easy implementation. In this work, P&O MPPT algorithm is employed to generate the reference voltage for sliding mode controller (SMC). The operartion of P&O algorithm is to perturb periodically the control variable and to compare the instantaneous PV powers before and after the perturbation. It generates the reference voltage V_{ref} after measurement of the panel power. If the measured power is greater than the previous power, the reference voltage is steadily incremented in the same proportion otherwise if not then it is decreased. Fig. 2 presents the P&O algorithm flow diagram and it is implemented in PSIM simulation tools.

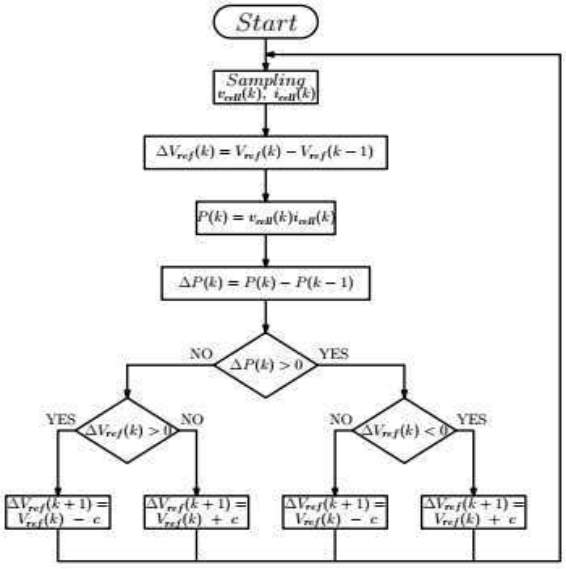


Fig. 2. P&O algorithm.

IV. Proposed Scheme

In this article an integer order SMC is proposed to extract the maximum power of the PV panel under varying sun intensity and temperature. Sliding mode controller is formulated here in order to control the proposed system [9].The objective of designing stable SMC for this system is to track the reference provided by P&O MPPT algorithm. The proposed scheme implemented in this work is presented in Fig. 3. In the implementation, at first the voltage and current of the PV array are provided to a P&O algorithm as input signals, which generates the reference signal which is further processed to the SMC controller. SMC controller performs the tracking of provided reference voltage signal and PV panel voltage signal

ISSN 1737-9334

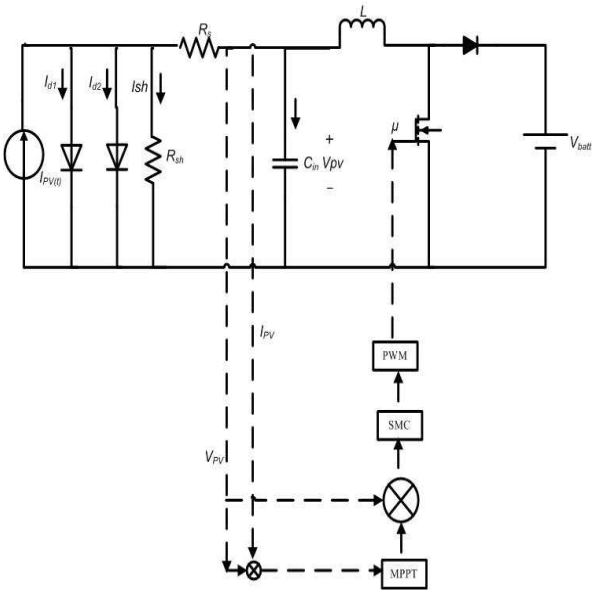


Fig. 3. System scheme based on improved SMC control.

to provide the control signal to the PWM of boost converter . The controller has been derived using the mathematical model of Boost converter and generates an output signal , which controls the duty ratio of the PWM signal provided to the converter switches. The dynamic behavior of dc-dc converter can be modeled as [10].

$$I_{cin} = C_{in}\frac{dV_{pv}}{dt} = I_{pv} - I_L \tag{2}$$

$$V_L = L\frac{dL}{dt} = V_{PV} - V_b(1-u) \tag{3}$$

Re arranging Eq. (2&3) we have

$$\frac{dV_{pv}}{dt} = \frac{-I_L}{C_{in}} + \frac{I_{pv}}{C_{in}} \tag{4}$$

$$\frac{dL}{dt} = \frac{V_{PV}}{L} - \frac{V_b}{L} + \frac{u}{L} \tag{5}$$

Eq. (4&5) can be written in state space form as

$$\begin{bmatrix} \dot{V_{PV}} \\ \dot{I_L} \end{bmatrix} = \begin{bmatrix} 0 & -\frac{1}{C_{in}} \\ \frac{1}{L} & 0 \end{bmatrix} \begin{bmatrix} \dot{V_{PV}} \\ \dot{I_L} \end{bmatrix} \tag{6}$$

$$+ \begin{bmatrix} \frac{I_{PV}}{C_{in}} & 0 \\ \frac{-V_b}{L} & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{-V_b}{L} \end{bmatrix} \tag{7}$$

Eq. 7 can be re-written in generalized form as

$$\dot{X} = f(X,t) + J(X,t) + h(X,t)u \tag{8}$$

Here $X = [x_1 \quad x_2]^T = [V_{PV} \quad I_L]^T$ represents PV panel voltage and inductor current $J = (X,t)$ and $h = (X,t)$ represents nominal system inputs and u represents the control excitation. I_{PV}, V_b, C_{in}, L represents PV current, bulk voltage, input capacitor capacitance and Inductor inductance respectively. The reference signal generated by MPPT algorithm can be written as

$$X_d = [x_d \; \dot{x_d}] = [V_{PV-ref} \; \dot{V_{PV-ref}}] \tag{9}$$

The tracking error can be defined as

$$\begin{cases} e_1 = x_1 - X_d \\ e_2 = \dot{e_1} \\ e_2 = x_2 - \dot{X_d} \\ \dot{e_2} = \dot{x_2} - \dot{X_d} \end{cases} \tag{10}$$

Sliding surface is chosen as

$$s = c_1 e_1 + e_2 \tag{11}$$

Here $c_1 > 0$ is the design parameter. Differentiating Eq. 11 on both sides

$$\dot{s} = c_1 e_2 + \dot{e}_2 \tag{12}$$

By combining Eq. 8 and Eq. 12 one can obtain

$$\dot{s} = c_1 e_2 + f(X,t) + J(X,t) + h(X,t)\,u - \ddot{x}_d \tag{13}$$

To derive the control set $\dot{s} = 0$,the desired response can be achieved by choosing the control law as

$$\begin{pmatrix} u = & u_{eq} + u_s \\ u_{eq} = & h(X,t)^{-1}\left[\ddot{x}_d - f(X,t) - J(X,t) - c_1 e_2\right] \\ u_s = & -h(X,t)^{-1} k_s sgn(s) \end{pmatrix} \tag{14}$$

Where u_{eq} is the equivalent control item to drive the nominal part of system , u_s is the robust switching control item. $sgn(.)$is the signum function defined as:

$$sgn(s) = \begin{cases} +1 & ifs > 0 \\ 0 & ifs = 0 \\ -1 & ifs < 0 \end{cases} \tag{15}$$

V. Simulation Results

PSIM and MATLAB/SIMULINK simulation platforms are used to perform the simulations of the proposed controller to verify its performance. The paramters of the photovoltaic panel are simulated with an irradiance step of 1000 W/m^2 in the instant 0 and after 50 ms the irradiance is decreased by 50%. The parameters are: open circuit voltage V_{OC} is 21.1 V, short circuit current I_{SC} is 3.8 A, and maximum power point voltage V_{mp} is 17.1 V. The maximum power point current I_m is 3.5 A and the temperature is 25C^o as a constant. The parameters of each Dc-Dc converter components in this paper are: capacitance valus is 22 μF, inductance value is 100 H, and an inductor current ripple of Δi_L = 1 A. Due to SMC, the converter switches to a variable frequency centered around approximately 49.2 kHz [9]. The MPPT algorithm has a perturbation period of 400 s, while the perturbation amplitude is 1 V for the P&O algorithm. The simulation results are shown in Fig. 4 and 5 shows the profile of PV panel voltage and reference voltage generated by P&O MPPT with perturbation aomplitude of 1 V. It can be seen panel voltage V_{pv} starts following reference generated voltage Vpv_ref once reached the steady-state after transient behavior. The integer

ISSN 1737-9334

order sliding mode controller tracks the maximum power point accurately compared with the conventional linear controller. Fig. 5 is the power time curve of the panel power and reference generated power by P&O. It is observed panel power P_{pv} and reference power Ppv_ref are following each other once reached the MPP with the SMC controller tracking.

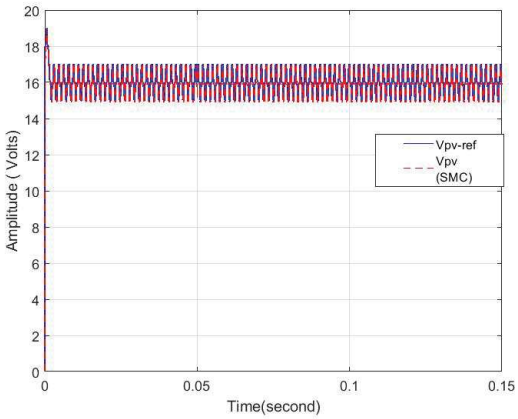


Fig. 4. Profile of the PV panel voltage.

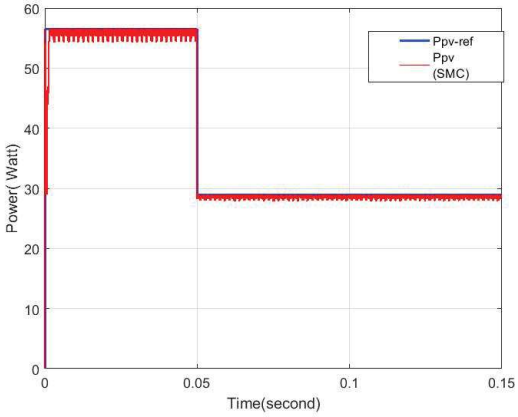


Fig. 5. Profile of the PV panel power extraction.

Fig. 6 shows the voltage tracking error response to study the behviour of SMC controller for the proposed PV system. It is highlighted that the proposed controller clearly outperforms the conventional controller in literature. It is highlighted that the designed system is robust to the environmental variation and has lower energy losses. It is also observed that the system efficiency has improved greatly when the proposed controller is used.

VI. Conclusions

In this paper an integer order sliding mode controller is derived and implemented for PV panels with boost converter. P&O MPPT technique is used to extract maximum power from the PV panel under varying irradiance condition. The controller part is implemented in Matlab Simulink environment

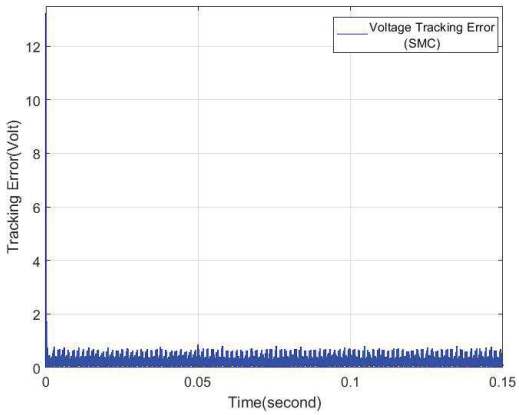


Fig. 6. Tracking response of proposed SMC controller .

and PV system along with MPPT algorithm is implemented in PSIM environment. The controller along with PV system is then co-simulated using Simcoupler . Response of SMC interms of voltage tracking and low power losses is superior in comparison with conventional controller. The simulation results show that SMC offers minimum transient time and overall power loss thus the efficiency of the PV system can be improved.

References

[1] P. Ekins, M. J. Bradshaw, and J. Watson, *Global energy: Issues, potentials, and policy implications*. Oxford University Press, 2015.
[2] M. Metry, M. B. Shadmand, R. S. Balog, and H. Abu-Rub, "Mppt of photovoltaic systems using sensorless current-based model predictive control," *IEEE Transactions on Industry Applications*, vol. 53, no. 2, pp. 1157–1167, 2016.
[3] N. Femia, G. Petrone, G. Spagnuolo, and M. Vitelli, "Optimization of perturb and observe maximum power point tracking method," *IEEE Transactions on Power Electronics*, vol. 20, no. 4, pp. 963–973, July 2005.
[4] D. Sera, L. Mathe, T. Kerekes, S. V. Spataru, and R. Teodorescu, "On the perturb-and-observe and incremental conductance mppt methods for pv systems," *IEEE journal of photovoltaics*, vol. 3, no. 3, pp. 1070–1078, 2013.
[5] B. N. Alajmi, K. H. Ahmed, S. J. Finney, and B. W. Williams, "Fuzzy-logic-control approach of a modified hill-climbing method for maximum power point in microgrid standalone photovoltaic system," *IEEE Transactions on power electronics*, vol. 26, no. 4, pp. 1022–1030, 2010.
[6] E. K. Anto, J. A. Asumadu, and P. Y. Okyere, "Pid control for improving p&o-mppt performance of a grid-connected solar pv system with ziegler-nichols tuning method," in *2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, 2016, pp. 1847–1852.
[7] J. Liu, *Sliding Mode Control Using MATLAB*. Academic Press, 2017.
[8] F. Attivissimo, A. Di Nisio, M. Savino, and M. Spadavecchia, "Uncertainty analysis in photovoltaic cell parameter estimation," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 5, pp. 1334–1342, 2012.
[9] D. G. Montoya, C. A. Ramos-Paja, and R. Giral, "Improved design of sliding-mode controllers based on the requirements of mppt techniques," *IEEE transactions on Power Electronics*, vol. 31, no. 1, pp. 235–247, 2015.
[10] D. G. Montoya, C. A. R. Paja, and R. Giral, "A new solution of maximum power point tracking based on sliding mode control," in *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society*, Nov 2013, pp. 8350–8355.

ISSN 1737-9334

Residual Networks for Malware Detection

Hoda EL MERABET

Faculty of Science, Abdelmalek Essaadi University
B.P. 2117 Quartier M'hanech II, Av. Palestine, Tetouan, Morocco
helmerabet@uae.ac.ma

***Abstract*— The research on malware detection field is strongly active, especially with the help of machine learning. Machine learning has been evolved into this field in the last two decades. Support Vector Machines SVMs, random forests, logistic regression and deep neural networks were employed in different combinations in order to get high performance results, and thus to correctly identify malware from benign files. Several researches considered dealing with executable files as images a good way to benefit from the advantages of Convolutional Neural Networks CNNs to detect malware. In this study we are using Residual Networks ResNets. These are a form of CNNs based on shortcut connections for feature reuse. Instead of providing the raw data to our model, we opted for choosing just the important features extracted from the PE header fields, and converting them to byte values, then feeding them to the ResNet as greyscale images. The obtained results were satisfying.**

***Keywords*— Convolutional Neural Networks; Residual Neural Networks; Deep Learning; Machine Learning; Malware Classification; Cyber Security**

I. INTRODUCTION

Recognition and identification of malware is an active area because of the huge number of malicious software registered every day. With the revolution of machine learning techniques, this field got more attraction to invest on it with the employment of these techniques. Traditional commercial anti-malware programs rely basically on the signature-based static technique, the technique that would never detect new malware as they use novel signatures. The ability of machine learning models to learn from training data, and subsequently to generalize to previously unseen data is a major reason to have an increasingly number of researches interested in this area. There are a lot of factors that influence the success of a model, particularly the types of extracted features, the feature selection techniques and the used machine learning algorithms [1]. In this paper, we chose to use the PE features to train our model and a residual network model for classification.

Deep learning models suffer from the well-known problem the vanishing gradient problem. This problem gets bigger when using a large number of layers. The gradient becomes too small hence the training doesn't happen efficiently. The apparition of residual networks ResNets for image recognition [7, 12] was innovative for providing a new solution to this problem. This kind of networks was also used in Natural Language Processing NLP [13], in morphological reinflection [14], in language identification [15], in semantic tagging and Part-of-Speech POS tagging [16] and in single image super resolution [17]. Inspired by the results of these studies, ResNets are used in our study for malware detection for the first time, to the best of our knowledge.

II. RELATED WORK ON MALWARE DETECTION

There are three approaches for identifying malware from benign files. The first one is the static approach. It relies on the extraction of signatures, which are values present in the files or values that should be calculated using a hash function for instance. The second approach is the dynamic approach. This one requires the execution of the files on a virtual environment or a sandbox to be able to detect the suspicious activities and visualize the behaviour of each file. The third approach is the heuristic approach based on machine learning, which has proven its success the last two decades. It can rely on static features, or on dynamic features or on both of them.

In [2], each sample from the training set was executed within a virtual environment to capture dynamic behaviour of it. The model was then built upon 300 vectors of nine collected activity metrics. These metrics are: CPU user use, CPU system use, RAM use, SWAP use, received packets, received bytes, sent packets, sent bytes and number of processes running. These metrics were afterwards transformed to 300 vectors of x-y coordinates using Self Organizing Feature Map SOFM. The dataset was constituted of 1188 PE files; 594 malicious and 594 benign. Half the files were used for training and the other half for testing. After taking the machine activity metrics each second in a 5-minutes time window, the researchers ended by having a dataset of 345,000 observations. For comparison, random forest was used as machine learning classifier in an experiment, and in another one logistic regression was used. The best accuracy, namely 86.70%, was obtained using logistic regression.

As for [3], API calls and operating system resource instances were used to identify API call graphs. These graphs were used as input features for the designed deep learning model. Multiple layers of Sparse Auto Encoders SAEs were employed to reduce the size of the original extracted features, followed by a Decision Tree classifier. The training dataset was constituted of 1760 samples, where 880 are malicious and 880 are benign files. 10-fold cross-validation technique was used for training and testing. The achieved detection precision was 98.6%.

The features used in [8] were n-gram system call sequences. A combination of two models was used, thus two

ISSN 1737-9334

pairs of probability values are got for each sample. After that they were compared to a threshold to get the final prediction result. The first model was the Long-Short-Term Memory LSTM. The information gain was used in order to remove redundant subsequences, then the LSTM model was fed, followed by a Max-Pooling layer then by a logistic regression classifier. The second model was the Random Forest model. Here the input features were API statistical features, which are got from the association of two API calls based on the comparison of the two API call sequences hashes. The experiments were conducted initially with each model separately. The results got by combining the two models were better. The obtained accuracy was 95.7%.

Since CNNs demonstrated high performances in image recognition, numerous malware detection researches took advantage of this type of algorithms. The work of [4] relied on a dynamic approach. The samples were executed in virtual machines, then process performance metrics were extracted. Each sample was represented as an image, the rows contained the executed processes, and the columns contained the 28 extracted features per process like the process status, the CPU usage percent and the number of written bytes. CNNs were employed to classify the samples. An accuracy of 90% was obtained on the testing dataset.

In [5], the raw bytes of each executable file were converted using an embedding layer, to be used as input features. The model was constructed from CNNs and Recurrent Neural Networks RNNs followed by a fully connected layer. A dataset of 400,000 files split evenly between benign and malicious files was used for training. And a distinct testing set of 77,349 files was used for the final test phase. An accuracy of 90.90% was obtained.

III. Deep Residual Learning

A. Problem with Deep Neural Networks

The intuition behind deep learning is that adding more layers makes the model learning more complex features. Accordingly, it becomes able to make more accurate decisions. However, [7] showed that continuing to go deeper affects negatively the performance of a traditional CNN model. They conducted experiments on the CIFAR-10 dataset [9] to classify images. Fig. 1, represented by the experiments of [7] depicts the performance decreasing going from a model with 20 layers to a model with 56 layers.

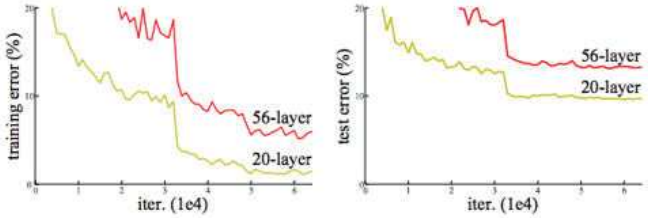


Fig. 1 Training error (left) and test error (right) on CIFAR-10 with 20-layer and 56-layer plain networks. The deeper network has higher training error and thus test error [7]

This problem of performance degradation can be returned to a very known problem, namely the vanishing gradient problem. The gradients become negligible when they pass from the end to the beginning of the model through the high number of layers. The problem can be caused as well by the loss of input information when it travels from the beginning of the network to its end. In order to lighten this problem while training very deep networks, the authors have introduced residual blocks. These blocks are based on feature reuse.

B. Residual Blocks

A residual block is a block of layers that adds the input to the output, before continuing to feed a new layer. It can be constituted of numerous convolutional layers, generally two or three but can be more.

Let us consider x the input to a few stacked layers of a deep machine learning model, and F(x) the output of it. If we assume that x and F(x) have the same dimensions, then a residual block can be defined by equation (1):

$$y = F(x) + x \quad (1)$$

Fig. 2 shows the building block of residual learning as given by [7].

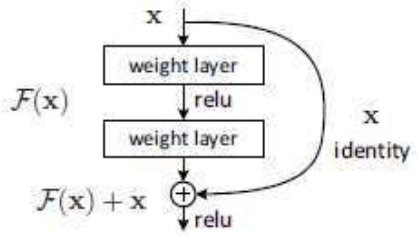


Fig. 2 A residual block [7]

The operation F + x is performed by a shortcut connection and element-wise addition [7]. It introduces neither extra parameters nor calculation complication.

If x and F(x) don't have the same dimensions, the identity mapping x can be multiplied by a linear projection just to match dimensions, as represented by equation (2):

$$y = F(x, \{W_i\}) + W_s x \quad (2)$$

Where: W_i are the weights to be learned in the stacked layers of the residual block

W_s is the matrix used to match dimensions

Residual networks are constituted by stacking consecutive residual blocks together. They use shortcut connections to propagate and reuse information over the layers of the model, the fact that allows us to build deeper networks.

C. Residual Networks in Image Recognition

In [7], ImageNet 2012 classification dataset [10] was used to test the efficiency of a residual network architecture. It consists of 1000 classes of images. The results are depicted in Fig. 3. In the left, training and test errors are drawn for two plain networks, the first one is constituted of 18 layers and the second one of 34 layers, without residual blocks. In the right, training and test errors are drawn for two residual networks,

ISSN 1737-9334

one with 18 layers and the other with 34 layers. We can see clearly that the validation error of the 34 layers-residual network is the lowest one among the validation errors of all other networks.

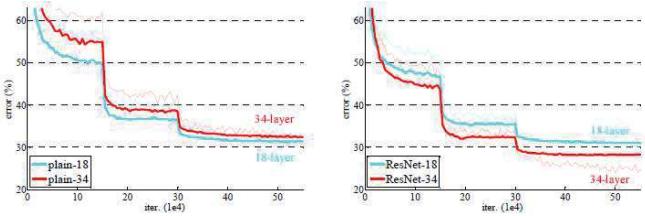


Fig. 3 Training on ImageNet. Thin curves denote training error, and bold curves denote validation error of the center crops. Left: plain networks of 18 and 34 layers. Right: ResNets of 18 and 34 layers. In this plot, the residual networks have no extra parameter compared to their plain counterparts [7]

In Fig. 3, the residual networks have no extra parameters compared to their plain counterparts.

The study of [7] demonstrates that by increasing the depth of the network, we obtain better validation accuracies. Adding more residual layers to the networks helps with generalization.

IV. Implementation and Experiment Results

A. Dataset Used

We empirically demonstrate the effectiveness of the residual network built in this study on the EMBER dataset [6]. It consists of 900K training samples (300K malicious, 300K benign, and 300K unlabeled) where only the benign and the malicious ones were utilized in our experiments. In addition to the training samples, it contains 200K test samples (100K malicious and 100K benign). The training of our model was done on 0.8% of the 600K samples training dataset, the remaining 0.2% of the data was used for the validation at the end of each epoch. The 200K test samples were kept to the end of the training just to evaluate the model on previously unseen data. The employed features are values extracted from the PE header fields. As shown in [1], various researches adopt the PE features for building their models obtaining high accuracy rates. Fig. 4 shows an example of the extracted features from an executable as given by [6].

B. Methodology

The vectorizing of raw features is provided by the EMBER authors [6], so each sample is represented as a feature vector of dimension 2351. A residual network needs images as input. To implement our malware detection model, we chose to represent each sample as a greyscale image of 50 by 50, so we padded each vector by 149 zeros, then we reshaped it to an array of 50 by 50 to feed the model.

CNNs were largely used for malware detection. As far as we know, residual networks were not yet employed in this field. Motivated by the success of these networks in image classification primarily, then in many other fields, we were convinced that they will give favourable results in the classification of benign and malicious files.

"**sha256**": "000185977be72c8b007ac347b73ceb1ba3e5e4dae4fe98d4f2ea92250f7f580e",
"**appeared**": "2017-01",
"**label**": -1,
"**general**": {
"**file_size**": 33334,
"**vsize**": 45056,
"**has_debug**": 0,
"**exports**": 0,
"**imports**": 41,
"**has_relocations**": 1,
"**has_resources**": 0,
"**has_signature**": 0,
"**has_tls**": 0,
"**symbols**": 0
},
"**header**": {
"**coff**": {
"**timestamp**": 1365446976,
"**machine**": "I386",
"**characteristics**": ["LARGE_ADDRESS_AWARE", ..., "EXECUTABLE_IMAGE"]
},
"**optional**": {
"**subsystem**": "WINDOWS_CUI",
"**dll_characteristics**": ["DYNAMIC_BASE", ..., "TERMINAL_SERVER_AWARE"],
"**magic**": "PE32",
"**major_image_version**": 1,
"**minor_image_version**": 2,
"**major_linker_version**": 11,
"**minor_linker_version**": 0,
"**major_operating_system_version**": 6,
"**minor_operating_system_version**": 0,
"**major_subsystem_version**": 6,
"**minor_subsystem_version**": 0,
"**sizeof_code**": 3584,
"**sizeof_headers**": 1024,
"**sizeof_heap_commit**": 4096
}
},
"**imports**": {
"KERNEL32.dll": ["GetTickCount"],
...
},
"**exports**": []
"**section**": {
"**entry**": ".text",
"**sections**": [
{
"**name**": ".text",
"**size**": 3584,
"**entropy**": 6.368472139761825,
"**vsize**": 3270,
"**props**": ["CNT_CODE", "MEM_EXECUTE", "MEM_READ"]
},
...
]
},
"**histogram**": [3818, 155, ..., 377],
"**byteentropy**": [0, 0, ... 2943],
"**strings**": {
"**numstrings**": 170,
"**avlength**": 8.170588235294117,
"**printabledist**": [15, ... 6],
"**printables**": 1389,
"**entropy**": 6.259255409240723,
"**paths**": 0,
"**urls**": 0,
"**registry**": 0,
"**MZ**": 1
},
}

Fig. 4 Raw features extracted from a PE file [6]

Various implementations were tested in our experiments. After comparison, we retained the implementation described below.

We chose to add noise to the input layer through a Gaussian noise layer. This layer has a regularizing impact and helps to lessen overfitting. After that, a 2D convolutional layer is added. 64 filters are used, which means trying to detect 64 different features. Then a Batch Normalization layer is added. As explained by [11], a batch normalization consists of performing the normalization for each training mini-batch. This procedure accelerates the training of deep neural networks. The chosen activation is Rectified Linear Unit ReLU. It is the more popular activation function nowadays. Next, a max pooling layer is introduced. This layer has no parameters to learn, it takes a region of the input image of a fixed filter size, then selects the maximum value of the different region values. Subsequently, two residual blocks are added. Each residual block consists of two convolutional

ISSN 1737-9334

layers with batch normalization and a ReLU activation function, followed by adding a shortcut connection to the output of these two convolutional layers, then passing the result through a ReLU activation. Finally, since we are involving binary decisions, either malicious or benign class, then we are using binary crossentropy as loss function.

C. Experiments and Evaluation Results

All the experiments in this paper were conducted on a laptop computer with Intel® Core (TM) i7-6500U @ 2.50 GHz, 2.59 GHz, and 16GB of RAM.

Based on the dataset described in Section IV.A, we evaluate the experiments in two perspectives; a residual network model without dropout regularization and a residual network model with dropout regularization.

After several attempts of parameter tuning, for each model type, the best results were obtained using a learning rate of 0.00007, a batch size of 256 and 64 filters for the different convolutional layers.

The figures 5 and 6 represent the training and validation loss and accuracy obtained with both models, without and with dropout regularization. After a certain epoch, the model continues to learn peculiarities of the training data and it is not any more able to generalize to previously unseen data. Therefore, we had to opt for an early stopping approach in order to get the best test accuracy. The epoch to stop at is different from a model to another.

For the residual network model without dropout regularization, the best validation accuracy was obtained at epoch 27, namely 91.01%, therefore we retained the model saved at this epoch. The test of this saved model on the 200,000 previously unseen test samples gave a test accuracy of 88.30%. The accuracy decreased by 2.71% from validation data to testing data. This decrease is a normal behavior of machine learning models as seen on [1].

As for the residual network model with dropout regularization, the best validation accuracy was obtained at epoch 79, namely 92.29%. At this epoch, the best test accuracy was obtained, that is 90.38%. Here we have an accuracy decrease of 1.91%.

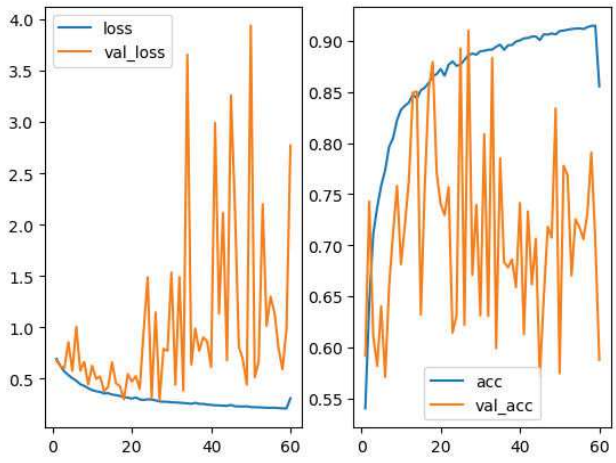


Fig. 5 Train and validation loss and accuracy obtained with our ResNet model without dropout regularization

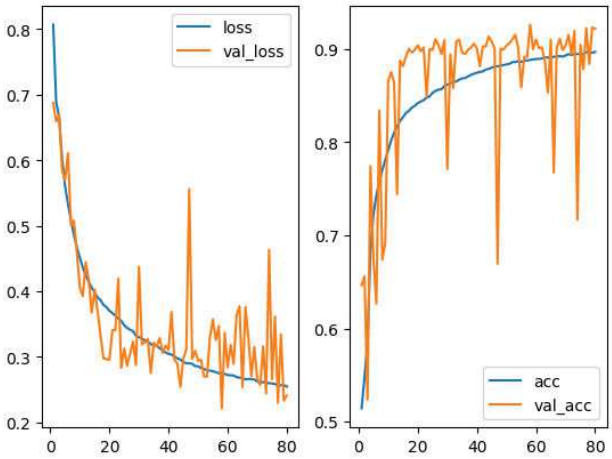


Fig. 6 Train and validation loss and accuracy obtained with our ResNet model with dropout regularization

The results of the different models are shown in Table1.

TABLE I
TEST ACCURACIES ON EMBER TEST DATASET FOR THE 4 STUDIED MODELS

	Our ResNet Model
Without Dropout	88.30%
With Dropout	90.38%

We see clearly that the residual network model with dropout regularization performs better giving us an accuracy of 90.38%. This accuracy is even better than some other previous researches as shown in Table 2.

TABLE II
COMPARISON OF OUR MODEL WITH SOME OTHER RESEARCHES

	Accuracy
Reference [2]	86.70%
Reference [4]	90%
Our ResNet model	**90.38%**

V. CONCLUSIONS

The use of residual networks in malware detection is new as far as we know. The model built in this paper gives satisfying results as shown. These results can be much more improved in the future with more hyper-parameters tuning. Implementing deeper residual networks may help on getting better results. However, the training of these networks can be painfully slow and needs more powerful machines with GPUs and larger RAM sizes.

REFERENCES

[1] H. El Merabet and A. Hajraoui, "A survey of malware detection techniques based on machine learning," *International Journal of Advanced Computer Science and Applications (IJACSA),* vol. 10, pp. 366–373, Jan. 2019.

[2] P. Burnap, R. French, F.Turner and K. Jones. "Malware classification using self-organizing feature maps and machine activity data," *Journal Computers and Security,* vol. 73, pp. 399–410, 2018.

ISSN 1737-9334

[3] F. Xiao, Z. Lin, Y. Sun and Y. Ma. "Malware detection based on deep learning behavior graphs," *Mathematical Problems in Engineering,* vol. 2019, Feb 2019.

[4] M. Abdelsalam, R. Krishnan, Y. Huang and R. Sandhu. "Malware detection in cloud infrastructures using convolutional neural networks," in *Proc. 11th IEEE Conference on Cloud Computing (CLOUD)*, San Francisco, CA, Jul 2-7 2018.

[5] E. Raff, J. Barker, J. Sylvester, R. Brandon, B. Catanzaro and C. Nicholas. "Malware detection by eating a whole exe," *AAAI Workshop on Artificial Intelligence for Cyber Security*, 2018.

[6] H. Anderson and P. Roth. "EMBER: An open dataset for training static PE malware machine learning models," in *ArXiv e-prints*. Apr 2018.

[7] K. He, X. Zhang, S. Ren and J. Sun. "Deep residual learning for image recognition," *Computer Vision and Pattern Recognition (CVPR)*, 2016.

[8] L. Xiaofeng, Z. Xiao, J. FangshuoY. Shengwei and S. Jing. "ASSCA: API based sequence and statistics features combined malware detection architecture" in *Procedia Computer Science 129,* Jan 2018, pp 248-256.

[9] A. Krizhevsky. "Learning multiple layers of features from tiny images," *Tech Report*, 2009.

[10] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein et al. "Imagenet large scale visual recognition challenge," in *ArXiv:1409.0575*, 2014.

[11] S. Ioffe and C. Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *CoRR*, vol. 1502.03167, 2015.

[12] K. He, X. Zhang, S. Ren and J. Sun. "Identity mappings in deep residual networks," in *ArXiv preprint* arXiv:1603.05027, 2016.

[13] A. Conneau, H. Schwenk, L. Barrault and Y. Lecun. "Very deep convolutional networks for natural language processing," in *ArXiv preprint*, arXiv:1606.01781, 2016.

[14] R. Ostling. "Morphological reinflection with convolutional neural networks," in Proc. Of the 2016 Meeting of SIGMORPHON, Berlin, Germany. *Association for Computational Linguistics*.

[15] J. Bjerva. "Byte based language identification with deep convolutional networks," in *ArXiv preprint* arXiv:1609.09004, 2016.

[16] J. Bjerva, B. Plank and J. Bos. "Semantic tagging with deep residual networks," in *Proc. Of COLING*, Osaka, Japan, Dec. 2016.

[17] B. Lim, S. Son, H. Kim, S. Nah and K. M. Lee. "Enhanced deep residual networks for single image super-resolution," in *the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* Workshops, 2017, pp. 136-144.

ISSN 1737-9334

Tracking Current and/or Potential Customers using Geofencing, Challenges and Open Issues

Keneilwe Zuva [#1], Queen Miriam Sello[*2], Tranos Zuva [#3]

[#*] *Department of Computer Science*
University of Botswana
Gaborone, Botswana
[1]zuvak@ub.ac.bw
[2]*sellomq@mopipi.ub.bw*

 # Department of ICT
Vaal University of Technology
Vanderbijlpark, South Africa
[3]tranosz@vut.ac.bw

Abstract—The **majority of people have become active mobile technology users wherever they are and anytime. They request information anytime, anywhere, so it has to be personalized and contextual to better satisfy their needs. At the same time the service providers would want to find out whether their clients are within their neighbourhood so that they can alert them of events or services in their proximity. This paper endeavours to highlight geofencing technologies, their uses and then look at the challenges and issues within this area. The building blocks of geofencing and the technology of marking the region of interest are explained. In conclusion it was foundout that geofencing can be used to offer personal and contextual services to mobile users anywhere and anytime.**

Keywords—geofencing technologies, contextual services, mobile technology

I. INTRODUCTION

The majority of people have become active mobile technology users wherever they are and anytime. Information is requested anytime, anywhere, therefore it has to be personalized and contextual to satisfy the needs of the users. A huge market has been created for new applications in the area of information relevance [5]. Location-based Services (LBSs) are one of the most ground-breaking applications that belong to these applications, they generate, compile, select, or filter information or perform other actions by taking into consideration the current location of the user [7]. Recently, social community platforms have also adopted the idea of LBSs to enable location sharing, that is, the mutual exchange of current whereabouts among their members [2].

It must be noted that location is just one example that describes the current condition/situation of the mobile user, there are many other parameters such as the time, weather conditions on the spot, the current means of transportation while travelling, or what the mobile user is doing [1]. Aggregate of all these parameters is taken into consideration for delivering the user with relevant information and there parameters are known as context. They are derived and processed by Context-aware Services (CASs). Thus, LBSs can be regarded as a special appearance of the more general CASs. The automatic detection of context parameters is still a technical barrier for a successful mass-market introduction of CASs. This is due to the fact that they often suffer from the non-availability of appropriate sensors in mobile devices or in the environment as well as from the potential variety of such parameters. The location of the user can be easily derived from various positioning technologies like GPS, Cell-ID or wifi. These technologies form the key for LBSs and their broad availability is one of the prerequisites for the latest commercial success of these services and are now standard features of smartphones and mobile networks.

The location of a user is determined by LBSs only when the user is active in a service session. Geofencing allows to detect users entering and leaving of geofences [8], which are pre-defined geographic areas. The user needs to be continuously tracked in the background even when the mobile device is idle or executing other applications all the time or from the time the user enters the region of interest depending on the type of geofencing available [6]. A broad range of new applications have been introduced due to geofencing and background tracking, especially in the area of information relevance. However, their introduction is also related to new barriers, such as the battery consumption on mobile devices has increased and users are now concerned about their privacy due to the availability of these new applications. This paper will look at geofencing, its challenges and open issues. The paper is arranged as follows: geofencing and background tracking in section two, challenges and open issue in section three and finally conclusion.

ISSN 1737-9334

II. GEOFENCING AND BACKGROUND TRACKING

Geofencing is determining the position of an object relative to a base position. It is a virtual perimeter for a geographic area in real world [4]. Thus, given a set of geographic areas G, each $g \in G$ is a polygon (P), which is described by a sequence of latitude (La) and longitude (Lo) pairs as shown:

$$P = ((x_1, y_1),(x_2, y_2),(x_3, y_3),.....,(x_n, y_n)) \quad (1)$$

where $x_i \in La, y_i \in Lo$,and $i = 1,2,3,....n$

These pairs follow some order, for example anti-clockwise, with its existing time period $[t_1, t_2]$. This virtual boundary/perimeter located a predetermined distance away from the base position is a geofence. We get spatial-temporal tuples O, each $o \in O$ is expressed into a form as:

$$(o_{id}, x, y, t) \quad (2)$$

where o_{id} is an object identified by id located at $(x, y)\, at\, time\, t$.

The important thing is to determine if the object being tracked is either inside, directly on, or outside the geofence boundary. Thus for each o_{id} and g , it is to find whether at any given time t and during the existence period $[t_1, t_2]$ the following conditions are fulfilled or not:

$$o_{id} \in g \; or \; o_{id} \notin g \quad (3)$$

In order to mark the virtual location of interest one has to use Global Positioning System (GPS) or a non-GPS which can also be termed satellite system or terrestrial system. In using the GPS one has to build an application that a user has to download to his/her device. When using non-GPS for geofencing one needs no application. The following section describes the two types of geofencing.

III. TYPES OF GEOFENCING

The types of types of geofencing to be described are Application based (App based), network based and hybrid geofencing shown in Fig. 1.

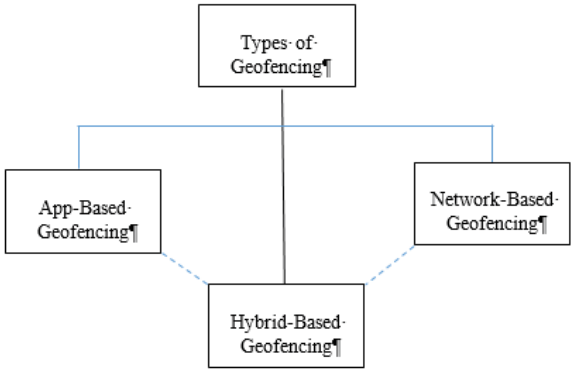


Fig 1: Types of Geofencing

A. *App-Based Geofencing*

App-based macro-geofenceing requires an application to access GPS data. In order to mark the region of interest one has to give the latitude and the longitude of the region in question then the radius. The radius is used to mark the virtual perimeter of the region of interest. This is shown in Fig 2.

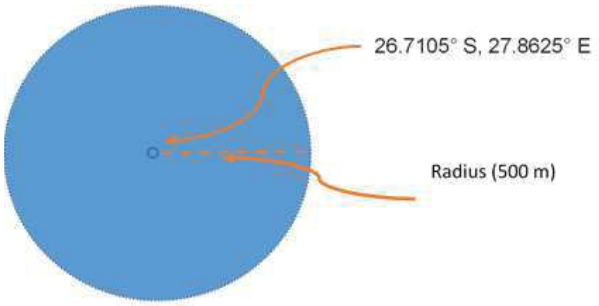


Fig 2: Geofencing of Vaal University of Technology

This is set at the server and then users can opt in to receive the messages related to what is being offered at the area of interest. Users must have a GPS receiver devices to be tracked and GPS data from the receiver is used to determine where the user is relative to the geofence. The goal of a geofence application is to report whether a GPS receiver is inside or outside the user specified geofence volume [9].

Geo-fencing technology enables remote monitoring of geographic areas surrounded by a virtual fence (geo-fence), and automatic detections when tracked mobile objects enter or exit these areas [4]. A Geo-fence becomes a layer of intelligence that allows users to make decisions or take some action based on a geographical area. It can circle anything/ any area you like

ISSN 1737-9334

– a retail store, a stadium, a neighbourhood. It must be noted that the user have to download the app on their mobile devices.

App-based micro-geofencing requires an application to derive the proximity to beacons. The beacons are the ones that form the geofence. The beacons can be placed around a building then the geofence is established as shown in Fig. 3.

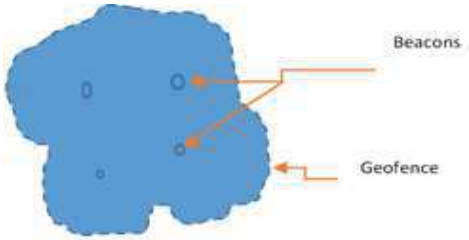


Fig 3: Beacon based geofencing

These devices whose range of transmission is approximately seventy metres (70m) repeatedly relay a discovery signal that is received by BLE enabled smartphones within the range of transmission. Beacons essentially allow a one-to-many experience.

Fig 4: One-to-Many mapping

As shown in Fig.1 one beacon transmits signals that are picked up by all the phones within the beacon's range. When a consumer's BLE enabled smartphone is within the range of the beacon, it receives and measures the signal strength in order to calculate the approximate distance from the beacon. The operating system of the phone then extracts the beacon ID and makes the ID available to the app on the phone. After these steps are completed, the app decides on the next course of action, based on the user's location.

B. Network based geofencing

A network based geofencing can be called a non-GPS based geofencing. A network based geofencing can be categorized into micro and macro –geofencing. The macro-geofencing would use the cell towers while micro-geofencing use wifi hotspots. With cell tower ID, mobile operator identifies the position of a mobile device through the ID of the Base Transceiver Station (BTS) that user connected at particular moment. The accuracy of the Cell ID method therefore depends on the known range of the particular BTS serving the user at the time of the query. It can range from a few hundred meters

ISSN 1737-9334

in urban areas to several kilometres (approximately one to two kilometres) in rural areas. So the macro-geofencing is marked by the BTS's strength as shown in Fig. 4.

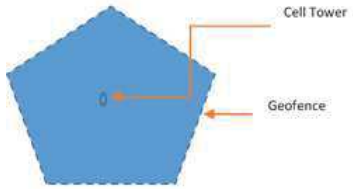


Fig 5: Cell Tower based geofencing

The wifi hotspots are used to mark the region of interest as shown in Fig. 5.

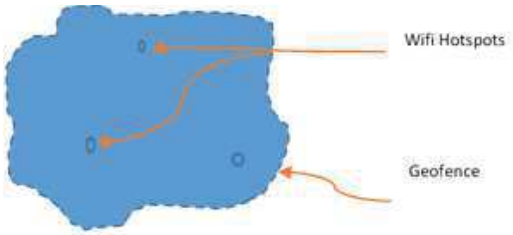


Fig 6: Wi-Fi hotspots based geofencing

Wi-Fi positioning, the general range of accuracy is at approximately 100 meters. Wi-Fi technology does not require the consent of the user because the user of wifi enabled device is the one who turns on the wifi on the device in search of free wifi networks. This is then used to detect the presence of the user who crosses the geofence into the region of interest. Wi-Fi is technically designed to accurately point to a device's exact location with wireless access points, by measuring parameters such as MAC address and SSID.

C. Selection of geofencing technology

In selecting the geofencing technology to use one needs to take so many things in consideration such as tracking the mobile user every time whether outside the region of interest or not, is it in doors or not, etc. Comparison of these technologies will help in deciding whether to use one technology or mix.

TABLE I: COMPARISON OF APP-BASED AND NETWORK-BASED GEOFENCING

	App-based Geofencing	**Network-based Geofencing**

Location Source	GPS	Bluetooth	Cell Towers	Wifi
Location range	As required	10 metres	Cell Tower range	100 metres
Accuracy of the object being tracked	30 – 100 metres	3 -10 metres	50 - 5000 metres	3 – 30 metres
Indoor Coverage	No	Yes	Yes	Yes
Battery Drain	High	Moderate	Low	Low
Internet	Yes	Yes	No	No
Download	Yes	Yes	No	No
Static Location	Yes	No	Yes	No
Service Location suitability	Out-door	In-door	Out-door and In-door	Out-door and In-door

D. *Hybrid Geofencing*

Hybrid geofencing is where the technologies for geofencing are combined to work together taking into consideration technologies that complement each other in the environment in question. The possible combinations is given in Fig. 7 below.

ISSN 1737-9334

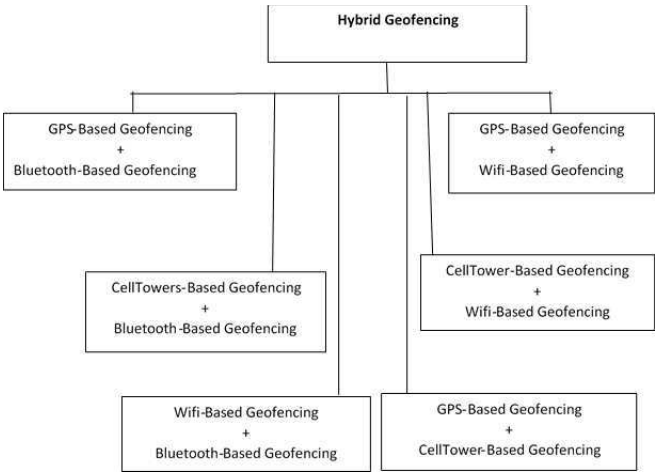


Fig 7: Forms of Hybrid Geofencing

In general the most appropriate combination is the macro-micro combination because it makes it possible to track a user the moment one enters the region of interest up to the point when the user is now on the proximity of the service provider. The micro-micro combination can be used to cover a small area of interest and can serve as a backup in the event of failure of one of the technology. The macro-macro combination can be used in difficulty places to go and where the region of interest is just temporal and there are buildings or caves that GPS cannot operate properly. The choice of the mix depends mostly on what is the objectives of the service provider.

IV. CHALLENGES AND OPEN ISSUES

Geofencing apps actually have to be running to know if a geo-fence has been crossed. This background tracking depletes mobile users' battery power. App-based Geo-fence only run on smartphones thus excluding a large number of mobile users who are using feature mobile phones. The macro app-based geofencing requires connection all the time thus consumes the mobile users' data and that translates into cost to the mobile user. App-based geofencing has added cost in building and maintaining the app.

The macro network based geofencing cannot pinpoint were the mobile user is within the region of interest. It is possible to track a mobile user without consent thus violating personal privacy of the mobile user.

Geofencing services deals with large heterogeneous data that need proper storage and analysis for it to be useful. Some of these applications belong to the category of Location-based Services (LBSs), which generate, compile, select, or filter information or perform other actions by taking into consideration the current location of the user [3].

REFERENCES

[1] J. Borràs, A. Moreno, & A. Valls, "Intelligent tourism recommender systems: A survey. Expert Systems with Applications", Vol. 41, Issue 16, 2014, pp.7370-7389

[2] F. Cirillo, T. Jacobs, M. Martin, & P. Szczytowski, "Large Scale Indexing of Geofences". Paper presented at the Fifth International Conference on Computing for Geospatial Research and Application, Washington, DC , 4-6 Aug., 2014

[3] A. Küpper, U. Bareth, & B. Freese, "Geofencing and Background Tracking – the next Features in Location-based Services", Paper presented at the Jahrestagung der Gesellschaft für Informatik, Berlin, 2014.

[4] S. Li, W. Sun, R. Song, Z. Shan, Z. Chen, & X. Zhang, "Quick Geo-Fencing Using Trajectory Partitioning and Boundary Simplification" Paper presented at the In Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2013.

[5] D. Martin, A. Alzua, & C. Lamsfus, "A Contextual Geofencing Mobile Tourism Service", In P. R. Law, P. M. Fuchs & P. F. Ricci (Eds.), Information and Communication Technologies in Tourism, 2011, pp. 191-202 Springer Vienna.

[6] S. Mazumdar, S. N. Wrigley, N. Ireson, & F.Ciravegna, "Geo-fence driven crowd-sourcing for Emergencies.", Paper presented at the Proceedings of the ISCRAM 2015 Kristiansand, May 24-27,2015.

[7] P. Patil, & L, K. V., " Survey of Travel Package Recommendation System.", International Journal of Science and Research (IJSR), Vol. 3, Issue 12, 2014, pp.1557-1561.

[8] F. RECLUS, & K. DROUARD, "Geofencing for Fleet & Freight Management", Paper presented at the IEEE, 20-22 Oct, Lille, 2009.

[9] D. Verechtchiagine, US Patent No., 2011.

ISSN 1737-9334

Editor biographie:

Dr. Ahmed Rhif (Tunisia) is a Researcher & Engineer (PhD, Eng). He has more than 11 years of experience on Scientific Research, Teaching and industrial projects. He is actually the Dean of the International Centre for Innovation & Development (ICID). Ahmed Rhif has worked as a Technical Responsible Chief in LEONI (International Leader of Wiring Fibers Companies) and has occuped also the task of Project Manager and Method Engineer in both SMSI (electronic development industry) and CABLITEC (Engineering automobile company). Then he was a Lecturer at both the Private University of Sousah (UPS) and the High Institute of Applied Sciences and Technologies of Sousah (ISSATso) and now he is working as Lecturer in the High Institute of Applied Sciences and Technologies of Al Qayrawan (ISSATk). His research interests include Modelling, Control Systems and Engineering as well as the implantation of the international standard of quality (ISO-TS 16949). Dr. Ahmed Rhif is also a Trainer of Trainers certified from the Canadian Centre of Training in Montreal and Coach certified NLP from the American Board of Neuro-linguistic programming.

Editor biographie:

Dr. Ahmed Rhif (Tunisia) is a Researcher & Engineer (PhD,Eng). He has more than 11 years of experience on Scientific Research, Teaching and industrial projects. He is actually the Dean of the International Centre for Innovation & Development (ICID). Ahmed Rhif has worked as a Technical Responsible Chief in LEONI (International Leader of Wiring Fibers Companies) and has occuped also the task of Project Manager and Method Engineer in both SMSI (electronic development industry) and CABLITEC (Engineering automobile company). Then he was a Lecturer at both the Private University of Sousah (UPS) and the High Institute of Applied Sciences and Technologies of Sousah (ISSATso) and now he is working as Lecturer in the High Institute of Applied Sciences and Technologies of Al Qayrawan (ISSATk). His research interestsinclude Modelling, Control Systems and Engineering as well as the implantation of the international standard of quality (ISO-TS 16949). Dr. Ahmed Rhif is also a Trainer of Trainers certified from the Canadian Centre of Training in Montreal and Coach certified NLP from the American Board of Neuro-linguistic programming.